7 604 G71 7 AD

UK Patent Application (19) GB (11) 2 125 409 A

- (21) Application No 8320975
- (22) Date of filing 3 Aug 1983
- (30) Priority data
- (31) 8222486 8312490
- (32) 4 Aug 1982 6 May 1983
- (33) United Kingdom (GB)
- (43) Application published 7 Mar 1984
- (51) INT CL³
 C12N 15/00 C07G 7/00
 C12N 1/00 5/00
- (52) Domestic classification C3H B7 C6Y 125 404 501 503 U1S 1296 1332 1337 2415 C3H
- (56) Documents cited Nature 1982, 299, 178-180
- (58) Field of search C3H
- (71) Applicant
 National Research
 Development Corporation
 (United Kingdom),
 101 Newington
 Causeway, London
 SE1 6BU
- (72) Inventors
 George Gow Browniee,
 Kong Hong Choo
- (74) Agent and/or Address for Service
 R. K. Percy,
 Patent Department,
 National Research
 Development
 Corporation, 101
 Newington Causeway,
 London SE1 6BU

(54) Genetic engineering

(57) It has been a problem to find an alternative, less time-consuming, and more reliable source of factor IX, a polypeptide which is essential to the human blood-clotting process and necessary for the treatment of patients with Christmas disease. In order to aid in the solution of the problem, there is provided recombinant DNA containing a DNA sequence occurring in the human factor IX genome, and includes recombinant DNA comprising substantially the whole sequence of human factor IX genome, which is

inserted in a cloning vehicle and transformed into a host, such as Escherichia coli. Other fragments of the sequence have also been cloned and the invention includes DNA molecules comprising part or all of the human factor IX DNA. There is also described cDNA derived from human factor IX RNA. Uses include the provision of an intermediate of value in the genetic engineering of a factor IX polypeptide precursor and thence manufacture of the factor IX polypeptide, and in making probes for use in diagnosing the presence of normal or abnormal factor IX DNA in patients with Christmas disease.

1st amino acid sequence: Glu-Cys-Trp-Cys-Gln-Ala

mRNA

5' GA_G^A UG_C^U UGG UG_C^U CA_G^A GCN 3'

Deoxyoligonucleotides synthesized :

3' CT_C^T AC_G^A ACC AC_G^A GTT CG (oligo N2A)

3' CT_C^T AC_G^A ACC AC_G^A GTC CG (oligo N2B)

2nd amino acid sequence: His-Met-Phe-Cys-Ala

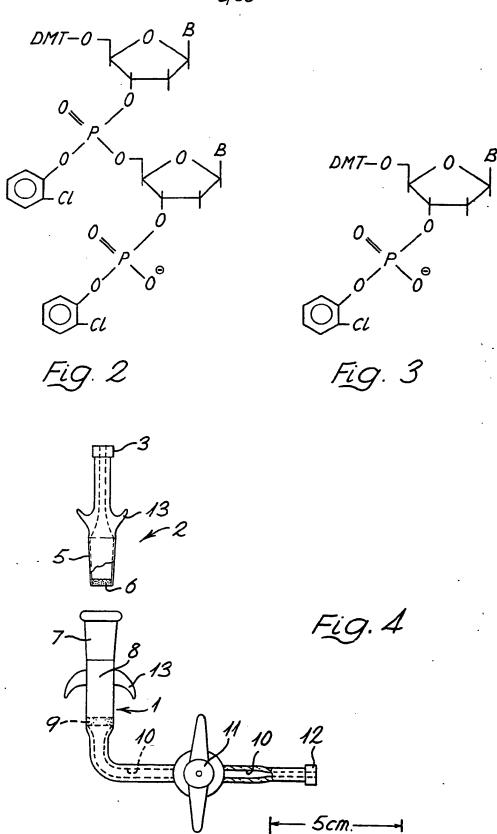
mRNA

5' CA_C^U AUG UU_C^U UG_C^U GCN

Deoxyoligonucleotides synthesized:

 GT_G^A TAC AA_G^A AC_G^A CG

(oligo N1)



ESNPCLNGGMCKDDINSY TGAATCCAATCCATGTTTAAATGGCGGCATGTGCAAGGATGACATTAATTCCTAT E C W C Q A G F E G T N C E L D A T C S I K <u>GAATGTTGGTGTCAAGC</u>TGGATTTGAAGGAACGAACTGTGAATTAGATGCAACATGCAGCATTAA N G R C K Q F C K R D T D N K V V C GAATGGCAGATGCAAGCAGTTTTGTAAAAGGGACACAGATAACAAGGTGGTTTGT S C T D G Y R L A E D Q K S C E P A V P F P TCCTGTACTGACGGATACCGACTTGCAGAAGACCAAAAGTCCTGCGAACCAGCAGTGCCATTTCC C G R V S V S H [V R P R F H G L C S C * E] CTGTGGACGAGTCTCTGTCTCACATGTGAGGCCCCGCTTTCACGGTCTGTGTTCGTGCTGAGAA 260 [270 280

Fig. 5

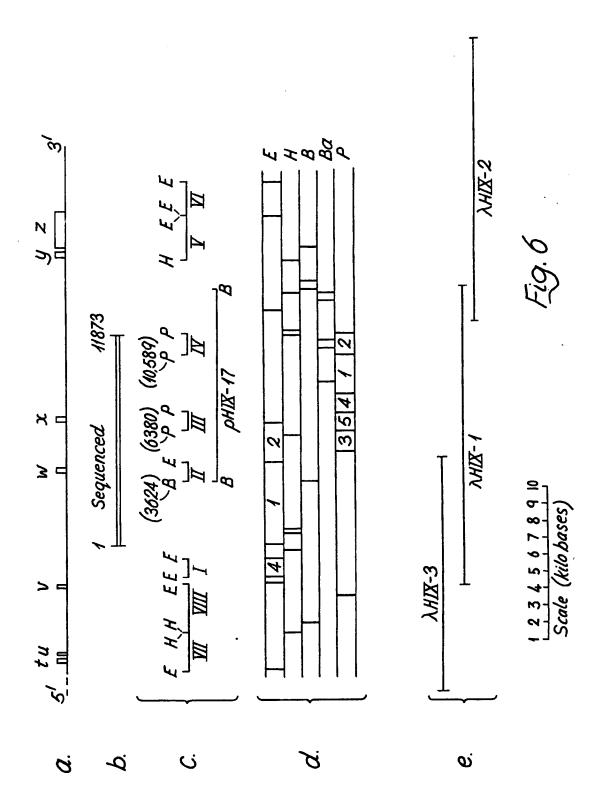


FIG. 7a

TCATTGGTTAGAWGTTCGACTTATGGGGAATTAACTC-CTCACATTTGCTAGTTGGTTGGGTACAGAGGGGT-ATAGTAGCAGCATTACTGCCTCA-GCATSAACAGGGAAGCTTTCA 970 , 980 , 980 1060 1070 1070 1080 1080 TCATSCCCTTASTGAATTATTGGTAGCAAAGGTTAAAGCTCAAGCTGSTTCCTTTGTCCCCTG-CAACAGTTGATTT-CCTCCCTTTATCTCCTGAAGTACCGTAAG-ACTAAGAGCCCAA 490 500 510 520 530 600 TTATTACATTTG-TCATGTCAGGCTATATGTAAAATAGAGTTTAAAAGTTTAGATTCACCTCAAAAATTCATATTCTCCAAAACCATACAGTCACTCTGTTAGCCTGTGT 819 020 630 640 700 710 720 CATTUTGC ATG CCCT USACA AACC AAGCT GCCATTTCG TA TCTCA ATGTTTG ACGAACACTTTC TACAGGTAATGTTTAGTTTGGCTGAACACTTTAGCAATTGCTTCTG S60 250 250 250 350 350 350 350 360 GAL TICCTISTGCCA ITATTITATIT CTGGAATCT TCASCCTT TTAGCTGASGGCAAAGATT GCTGATTAGGAAGCAATATTTCCCACCTCCTGCGCAAAACAAGCCAAAGATCAACAG

				6/3	5			
46676446C 1200	TTGTCAAAC 1320	ACAAGTCCT 1440	1560 1560	36CCAAAAT 1680	1800	1920	CATGTGAA 2040	CCCACCCT 2160
AAAGACACA	STTTTACTG	ICTAAGCTA! 1430	ICAGGAAACI 1550	16ACTGGGA(IAGACAAGAG 1790	CCCTATITC 1910	TTAGGGGTA 2030	CTGCCTCAT 2150
CATGCTGGA. 1180	CTATTGAAA(1300	TTCAGCAGG	CCAAAAATA) 1540	4 A T A A A A T T /	ICATTICAA!	1900 1900	FATTT FAGG1 2020	.TCTGCTCCT
AATCACCAC 1170	TGAGGATTT 1290	1410	AGGAATAAG(1530	GATACCTAA. 1650	AGAGGAAAA' 1770	CTTCAGTTA(1890	7 A A A T G T T T 1	46TG4TCTT1 2130
1160	CACTGACCAC 1280	.AGAGCCTGT 1400	CAAGATGGC 1520	AACGAAAGT 1640	ATAGATTCA 1760	GTCTTTGGT(1880	2000	STACCAAAT! 2120
TTGACATCC4 1150	1270 1270	SGCTCAAGGC 1390	1513	STATACTGTT 1630	NTCAGACAAA 1750	GACATATTG 1970	1990	TTCAGCTCA 2110
66A66CT 66	CTTTCCA 16 [°] 1250	AC AGAAA TA(1380	5 T A A T T A GA / 1 5 0 0	4 A T T T A A T T (1 6 2 0	1 7 4 0	7 A G C A C T A A 1 1 8 6 0	.TTCCTA GCA	SACGCAGATA 100
ACACCCATA 130	ATTTTCAAA 250	GCTCTAGAA 370	GAAAGCACA 490	GGAGTTAAA. o1û	730 730	TGGAATSCT1 450	CCTAATGGAG 970	TTATTACATO
TTCAAGGAT 20	CAGTCACTC 40 1	rcc	AAACAAAGG 80	TTTCAGCTT 00 1	AAAAGAAAC 20 1	57666CCAC	59 1	7161ACATA 30 20
CC & A G G C A A 0 11	ТАТСТА 4 TA 0	TGTTTATST 0	6CATAGGAG 0 14	AATGACAAC O 16	AAACTAACA. 0 17	1445444T(TTCAAGTTG(16666TTTG1
AGCTATGAG	CATAGGAGG 123	GTCTTACAT 135	ATTGGTATT 147	TCTATTAAT 159	aTCAGGT34,	12664645T,	1950 1950	atgtgtcac/ 207(
. ACAT AGTGC 1100	. A A3 G T G G T G	.665 A AA66 T	CTTCCTGTT 1460	TTAAACTAA 1580	aaaa Gaac T 1703	GGAATTGCA 1320	GTAACCHCC 1940	TAGA TAAACI 2060
AG G CAAAAGAC ACATAGTG CGCCTATGASCCAAGTCCAAGGATACAC CCCTAGGAGGCT GGTTGACACCCCCCCGGGGCTAATCACCCACCATGCTGGAAAAAGACACAGGTGAAGC 1030 1100 1110 1120 1130 1140 1150 1150 1200	TGAGAAGAATGAASGTGGTGCATAGGAGGTATCTAATACAGTCACTCATTTTCAAACTTTCCATGTTATGATTGCACTGAGGATTTCTATTGAAAGTTTTACTGTTGTCAAAC 1210 1220 1230 1240 1250 1250 1250 1250 1270 1280 1290 1390 1300 1300	ACGTACACAAGGGGA AAGGTGTCTTACATTGTTTCTTA TGTTCCTGCTCTAGAAACAGAAA TAGGCTCAAGGCAGAGCCTGTTTTTCTTAATTCAGCAGGTCTAAGCTAACAAGTCCT 1330 1340 1350 1350 1250 1370 1380 1390 1400 1410 1420 1430 1440	GAAACATGSTACTTCCTGTTATTGGTATTGCATAGGGAAACAAAGGGAAAGCACAGTAATTAGAAAATACAAAAACAAGATGGCAGGAATAAGCCCAAAAATATCAGGAAACACAATTATT 1450 1460 1470 1480 1490 1500 1500 1560	GTGAATTGGGATTAAACTAATCAATTAATAATGACAACTTTCAGCTTGGAGTTAAAAATTTAATTGTATACTGTTAACGAAAGTGATACCTAAAAATTACACTGGGAGGCCAAAAT 1570 1580 1590 1670 1680 1670 1680 1680 1680 1680 1680 1680 1680 168	GüA GGGA TGTG AAAA GAAC TATCA GGTAAAAAA CAAAA AAGCAAT CTTA ATATCA GACAAAA TAGATTCAA GAGAAAAT TTTCAAAA GACAA GA 1690 1703 1780 1780 1790 1800	TATTAATAAGGGGAATTGCATAGGAGAGTAAASAAAGTGTGGGCCACTGGAATGCTTAGCACT AATGACATATTGGTCTTTCAGTTACCTTACAGGACCCTATTTCATTCTCTT 1810 1320 1830 1930 1910 1950 1950 1960 1920	ATGITTGATATGTAACC…CCTCAGCCAGCTTCAAGTTGCTTTTTTGGCCCTAATGGACTTCCTAGCACTATAATTTCTTTTTTTAAATGTTTTAGGGTTTAGGGGTACATGTGAA 1930 1940 1950 1950 1960 1970 1980 1990 2000 2010 2020 2030 2040	GGTTTGTTACATASATAAACATGTGTCACAGGGGTTTGTTGTACATATTACATGACGCAGATATTCAGCTCASTACCAAATAGTGATCTTTTCTGCTCCTCTGCCTCATCCCACCCT 2050 2060 2060 2070 2080 2090 2100 2110 2120 2130 2140 2160
⋖	-	ব	ق	ق	ق	Ë	7	હ

IGACAGTGTGGTACTGACGATAGACAAATAGATCAGTGAAACACACTAGAGTGCTCCAGAAGAAGCACCTGTACATA 160 3270 3230 3240 160 3170 3180 3290 3200 3210 3220 3230 3240 170 3290 3390 3310 3320 3330 3340 572	TIT GGTGGTGG GAGGTCGGGCAGGATTCATAAAGTTATAAAATGCATATGCAAAGAGCCAAGGATAGCCAAGACATTTGAGGAAGAATAAACTTGTACTTACACTA 3010 3020 3030 3040 3040 3050 3060 3060 3070 3080 3090 3100 3120	CAATTAGAGGATACAATATAGTAATATCAAATGTTTACTAATGGATTCAATGCAATACCAAASGTCCCAGCAGGCTTTT 120 293U 2940 2950 2960 2970 2980 2990 3000		2860 2970 2980	G A G A C G T T A T T G A G A C B A C A C A C A C A C A C A C A C A	TITCAGAABAAATGATTSTACATATAGAAAACCCAAAGCATCTAAACTAAA	AATGCAGAAAGGCAGTATCTACATGAGATTATGAAATTGCGGTTGCTTTTTGTGTTCACTGAAAAAAAA	2500 2510 2520 2500 2510 2520 2620 2630 2640 2620 2740 2750 2760 2740 2750 2760 2750 2760 2760 2770 2760	CAAAAGGCATAATTCTTTTTTTTTTTTTTTTTTTTTTTT
ACACTAGAGTG 3210 14GATATTCAT 3330	677776866886 3090		GATTCAATGCA4 2970	2850 Gattcaatgca4 2970	CCTACATAAAAA 2850 Gattcaatgca ⁶ 2970	GAGTCAACATA 2730 CTACATAAAAA 2850 GATTCAATGCAA	2510 2510 2730 2730 2730 2850 2850 2970	2490 2490 115161164611 2510 2730 2730 2850 2850 2970	6614141646 2370 2490 2490 2510 2730 2730 2850 2850 2970
ATCAGTGAAAC 3200 GATAGGTCAAT 3320	TAGCCAAGACA 3080	GTTTACTAATG 2960		2840	1GTSCAAGAAC1 2840	TACTGGATACA 2720 1GTGCAAGAAC1	GCGGTTGCTTT 2600 17aCTGGATACA 2720 1GTSCAAGAAC1	CTTTTATAGGT 2480 2600 TACTGGATACA 2720 1GTGCAAGAAC1	TAATATTCCATI 2360 CTTTTATAGGT 2480 2600 TACTGGATACA 2720 1GTSCAAGAAC1
TAGACAATAG 3190 CATAAAAGT 3310	а GA G C C A A G G A 3070	ATATATCAAATI 2950	,	2830	GT G A G A A G A T A G	TATAGAAAGAT 2710 GTGAGAAAGA1 2830	41141GAAA1T 2590 1414GAAAGAT 2710 GTGAGAAGA1	TCCAGATTCTA 2470 2590 TATAGAAGAT 2710 GTGAGAAGA1	774766CTGCA' 2350 TCCAGATTCTA 2470 ATTATGAAATT 2590 TATAGAAGAT 2710 GTGAGAAGA1
5ACACA456A' 3180 1166161111	SGAAATGCAA! 3069	LATATAGTAA 0 2940	2820		астаатаарт	aetaaataa6 2700 2700 asteataaat	TCTACATGAG 2530 AATAAATAAG 2700 ASTAATAAAT	TCTAAAGACT 2460 TCTACATGAG 2530 AATAAATAAG 2700 AGTAATAAAT	CCCTTCTTTT (2340 CCTAAAGACT 2460 TCTACATGAG 2530 2700 2700 ASTAATAAAT
61616614C1 3170 3170 446644114 3290	AAATGCATATI 3050	TAAAGGATACI 293U	2 2 2	C 7	TT 26 AC GC TT	AAACAATTAA 2690 2690 11125ACSCTT	AAAGGCAGTA 2570 AAACAATTAA 2090 TTAGACGCTT	CACT GT AATT 2 450 AAAG GC AG TA 2 570 2 690 111 15 AC 5C 11	AGACATAATC 2330 CACTUTAATT 2450 AAAGGCAGTA 2570 AAACAATTAA 2090 TTAGACGCTT
at taagaca 3160 3160 3260 3280	.CTAATTATA. 3040	TT TATCAAT'	7.003	00 & 0	3C ATT AAAA	2680 2680 36 att maaaa 3000	TCAATGCAG 2560 .AAAGCATCT 2680 .CATTAAAA	CTATTCTAA 2440 TCAATGCAG 2560 SAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	TCCCACAAA 2320 CTATTCTAA 240 TCAATGCAG 2560 AAAGCATCT 2680 3CATTAAAAA
.aGTTACATT1 3150 150 146GTGCCAG1	GATTCATAAG 3030	ATAGGCAATG 2010	0642			17 16 16 16 16 16 16 16 16 16 16 16 16 16	AAATTACACI 2550 ATAGAAAACCC 2670 TTTTATAATA	TCCATGTCTG 2430 4444TACAC1 2550 173GAAAACCC 2670 171TATAAATA	CCATTCATAT 2213 2439 2439 2550 2550 2570 2670
CCAGATGTCAAGACTTATTATCGAGTTACATTTATTAA 3130 3140 3150 31 1140 3150 3140 3270 33	667CGGGC&G 3020	TTA AGGA A A A C CTA A A T A A T G A T G T T T A T 2 9 9 9 2 9 9 9 2 9 9 9 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	2780		AC C A 3 C A A C G A T T C A A A A A 1 G A T T T T T A T A T A T A S C A T T	ATTGT ACATA 200C 200C 20 A A A A A A A A A A A A A A A A A A A	ATTICTAGAATAACTAAGCAATAGAAATTACACTTCAA1 2530 2540 2550 256 TTTCAGAAAAATGATTGTACATATAGAAAACCCAAAGG 2650 2670 2670 2670	TC	TCC TAAGS ATG ATAGCCTCCAGCTCCATTCATATTCCCA 2230 2300 2310 2320 2420 2420 2430 2430 2440 2420 242
161CAAGa 3130 3130 46GCTTGa 3250	3010 SA	GAAAACCT. 2990	2770		CAACGATT	GBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB	7 A G A A T A A C 2 S 3 D 6 A A A A A T G 2 S D C A A C C A T C C C A T C C	541666647 2410 14644746 2539 64444476 2650 2650	166 a T 6 a T 6 a 7 a 6 a 7 a 6 a 7 a 6 a 7 a 6 a 7 a 6 a 8 a 7 c 6 a 8 a 7 c 2 6 5 0 a 7 c 2 a 6 c 4 T c 6 a 8 c 7 c 6 a 7 c

T 4330 4220 4230 4230 4250 4250 4250 4270 4280 4290 4300 4310 4320 4320 4210 4220 4220 4230 4230 4250 4250 4250 4250 4250 4250 4250 425
CATTIAGITITAAACCAATCAATTATAGIGCIACCAICATITITAIGGAAGITITATTITACCTITCTTCCACTCTTATTICAAGGCTCCAAAATTTCTCCCCAACGIA 4210 4220 4230 4230 4240 4250 4260 4270 4280 4290 4290 4300 4310 4320
TCATTTACTTTSTGAMAACTTACACTAAMATTGTGTGTATTTTTTGAATATATGTTTACATT AATAAATAGGGTTTTTTAAACCTGTAGTTCATAATTTAGTGAAAGTAGAATATCCAAA 4000 4100 4110 4110 4120 4130 4140 4150 4150 4150 4160
GTTASAAGTCC AGUTAATGGTAACCTATAAAAAGGAAAAAGGGTGGAATSATTSGGAGGGGGC ATCTTCTGGGGTATTGATAATGTTATTGGTCAGTTTAAACAGGC 3970 3980 4060 4070 4070 4070 4080
ACT GCT GTG TAAC AAC AAC AAC AG AG AG AG AG AG AG AG AG AGT AC AG AGT ACT TO TO TO TO AGA AGA AGA AGA AGA AGA T 3850 3860 3860 3870 3890 3890 3900 3910 3920 3930 3960
Cactattatacataa sagccaaaaactsgaa accaaatatccattaa cagtagaa tsaa taaataa agctgtaa tagtaggaa tactacaggaatgtaaatgaact 3730 3740 3820 3830 3830 3840
aaattgettgeeagtaatetagatetgaacatgigateeagtaattaeaeteatatataageeagtaaaaggeatgittatgieaeeagaagatataeaagaatgiteatta 3610 3620 3530 3640 3550 3660 3670 3680 3690 3700 3700 3700 3700
acactacticgactgaaaagacaagtcacagagtgagacaagatatcigcaatacagatacctaataactgaacccatacagtggtgggaatttaagttcgtacaatcattttaga 3490 3500 3500 3510 3529 3530 3540 3550 3560 3560 3570 3580 3580
aatitatatatataaditigcagaaagcaaaaatticttaaaatacaaaagssaatcaccataaaggaaaagattgataaactgactatataaaactaaggactcctgttcagcaaaag 3370 3360 3360 340 3470 3470 3480

				.,					
GAACCASCASETCATAATCTGAATAAGATTTTTTAAAGAAAATCTGTATCTGAACTTCAGCATTTTAACAAACCTACATAATTTTAATTCCTACTTGGATCTGCTTCCTTTTGAAATCA 680 4870 4880 4890 4800 4800 4810 4820 4830 4840 4850 4850 4880	TAGAAAATATCAGTAGCTTGAATTAGACCAATTAATTTTCTAGATTGCATCATATTTTAAATATAAACTATGTAATCATCTGAACTGAATTCHTTCTGAGTCCAATTTGTCCAATTTT 4890 4700 4700 4710 4720 4730 4740 4750 4760 4760 4770 4780 4790 4800	TITCTCTAACATTTATATCACAAATTAATTTGTGTGATTTCTGCATATGTATTTGTAATTCATCAAGTCAAATGTAGTAATACTATATCATAAAAAAAA	GTGATAGGCTTCTAGTATAAGGACGGTAAGTTTGAAGCATGATTCTATCTGGCTAGTTTACTCTGAGAAAGTTATTTTTTTT	GTCAGAATGATTCCGGCAATGAACTGTTTTATGTTCTGCTAGGCTGATCAATCTATATGGCTGTGAACAAACA	TGATTAGTSTATTCAGAACATCTCCCACTCCCATGTTCGTATGCTGTTGAAGCCAGTAGACACTTTTTATTTTGAAAATTTAGGCTCTGCAGGGTCAATTATATTTGAT 5170 5160 5160 5200 5210 5220 5220 5230 5240 5250	AAATGAGGGSCTITITILAAGGAAACTAGATATAATTICTITIGCATTICIAAAGCCTGATATCITATTAATTGGTACATTAAATTGTGCACCATTICTCTGTAACTGTTTCAGTACCTG 5290 5300 5310 5320 5330 5330 5340 5350	TCT CAGCACTATACCAGGCAGAAGAAATTAAAGAAA AAGAACCAGTGC CGAGATCAGGCAGGGAGGCCCTAATCCTGCGGCACTAGAGGAATTAAAGACACACAC	GASTATGAAGTGSGAAATCAGGGGTCTCACAGCCTTCAGAGCCTGAGAGCCCCGAACAGATTTACCCACATATTTGACAGCCAGC	TGGGAAATAAAGGATGAGTCSGCTAGTTATCTGCAGGAACATGTCCTTAAGGCACAAATCACTTATGCAATTGTCTGTGGTTTAAGAACACCTTTAAGCAGTTTTCCGCCCTGGGT 5650 5660 5660 5660 5670 5680 5690 5700 5710 5720 5730 5740 FTG. 78

FIG. 7f

				10/3	5				
.6666CCA 5880	6000 to 1	65C4TACT 6120	36CCAAAC 6240	ACTTGCTG 6360	CAAATTCC	6480	CAATCTCT 6600	1441CA14 6720	AACACAAA 6840
36CCAGTTT1 5870	1611C11C1C 5990	TAGACAGAT(6110	6 A A T A A G T A (6 2 3 0	TTCTCAACC 6350	GTTTSACTA	0449	. TT A T A T T A G 6590	AAACCTTCT 6710	. A A S A A A C T A 6 8 3 0
111G111A1	CTGCCCACT 5980	CTTCCAGTA 6100	GCTTTAGCA 6220	1775CTGTT 6340	AAGTGACTTT	6460	6CTGTGATAA 6580	TCTGCAAAAG 6700	CTTTTTAST. 6820
CTGCAGAGAT 5850	66CTTC66GC 5970	aaaaacctaa 6090	TATAGCCTT4 6210	TTTGAACTA/ 6330	TSAAATGT6/	9450	616441416 6570	5GCCTTTTAG 6590	SGSAGTAAGA 6810
.ATCACAGT6(5940	SAATCTCTAA S960	TTTTAACCTT 6080	6200 6200	ACATAAGTAT 6320	TAAATGAAAA	0779	GCAGGTCCTG 6560	AATCTAGAA(6680	ATGGCAATT(6800
TCACGAGCA1 S§30	TAAAATGAA(S950	6TCT5T6TA1 6070	ATTAAGTGT: 6190	TACACTTAT	IGG A A A G T G A	5430	1 A A A T G T G G G G G G G G G G G G G G	16C4CCTTGG 5570	CCATTCAGGG 6790
17CAAGGCCA 5820	ATCCTGCA AA 5940	3TGGGTGT ST 0060	ACCTCCAACT 6180	6TTGATTG TG 6300	GTGGCTTG AA	6420	ACCCAAGGA/ 6540	TATATAGA' 6660	ATCAGACT TO 6789
GGGCCALGTSTTCCTTGCCCTCATTCTSGTAAACCCACAAGCTTCCAGTGTSGATATCAAGGCCATCACGAGCATATCACGAGGATTTTGTTATSGCCAGTTTTGGGGCCC 5770 5780 5780 5860 5870 5870 5870	GTTTATGGCCAGATTTGGAGGCCAACAAACCAGAAGCTAGGAATATATACCTGCA AATAAAATGAAGAATCTCTAAGGCTTCGGGCCTGCCCACTTGTTCTTCTGCCTGGTT 5390 5900 5900 5910 5990 5990 6000	GGAGCATGAATATGTGTGTGTGTGTGTGTGTATTTTAACCTTAAAAACCTAACTTCCAGTATAGACAGATGGCATACT 140 6050 0060 6070 6080 6090 0100 6120	.aaacagaattgagaaccacctccaactattaagtgttatatttgaatatagccttagctttagcagaataagtaggccaaac 60 6170 6230 6230 6240	TTA A A A TA A SECTITICACE TO SECTIVE TO THE TOTACACT TO THE T	IT STCT GTTACAA GGAATGT GGCTT G AA GGAA A GTGATAAAT GAAAA TSAAATGT GAAGT GACTTTGTTT SACTACAAATT CC	6413	CATTCTSSTASTCCC CASTSTATCAATACTTTATTTTTTTTTT	TTG GCT A STATTTG A SCCCA A CHART CACA A SATOTTC CCCAGA A A STATATA A SOCCTTG GAATCTAGA A GGCCTTTTA GTCTGCA A A GAAA CCTTCTTA A TCACATA A SOCCTTTTA GTCT GAAA GAAA GAAA GCTTCTTA A TCACATA A SOCCTTTTA GTCTGCA A A SOCCTTCTTA A SOCCTTTTA GTCT GAAA GAAA GAAA GAAA GAAA GA	4GC 1GC 1GC 1GC 1GC 1TT 1 ACCAAATT 1 GAAA GOTT ACAAA GCATCATCAGACT TCCATTCAGGGATGGCAATTGGGAGTAAGACTTTTTAGTAAAGAAACTAAACACAAA 5730 6740 5750 6810 6830 6830 6840
ACA AÇCT TCC 5800	CCAGAAGCTA 5°20	GAGSAGC ATÚ 6040	AGAA4CAGA; 6160	.TCCCTTTTC1 5280	.AATGTETĢT	9400	111CTTTAGA 6520	AATSATCTCT 5640	aabsttacaa 6750
5GT&AACC.C.	CCCAACAAA 910	CTTCACATACACTGTUTCAAAGCTAGTCTACCTTGAGA 6910 5529 6939 69	AGCTAAACCCTTACAAGTTCTTCTATGCTATAAAAGAG olso olws 6150	STGATAAGG 5270	TCTTC4TS4TACTTT4TC6C4SCT5GTTGCT4TAGAAA	98290	TACATTATT1 0519	TTGAATCAC 1 6639	TSGAAGTT (5750
CCCTCATTCT 780 5	64 GGCCTGTT 90 G	TCAAAGCTAG1 5520	GIICIICIAIC 014G	GCCTTTTCA/	.SécaSCTés?	638û	STSTATCAA SOU	CCCAAATAA oolu	ITT ACCABAT 574 Ü
15 T CC T T CC 7 7 57	CCA 5ATT TGC 90 58	ACA CTGT LT (10 of	CCTTACA AG'	45C TTT CT(50 0.	7	9 0259	746 TCCC CA 30 5	178 T T T T G A S 5 10	SAAGTCCCAT
966CCA1G	611TATGS(53°	CTT CACATACA 6910	AG C T AA A C C C 0 1 30	TT & & & & TT & & & & & & & & & & & & &	TCTTC4TS	P) 9	CATTCT65	TTG6CT44	46C 46C 46

78
FIG.

			11/35						
SAAAATAACSCAATCAACCTTTAGCTTGAGACTCTATTCACTGATTAGATTTTTTTAAATACTGAGGCCTGCTTCTCAGAAGGAGGGGGGGG	V P F P C G R V S V S Q T S K L T R A E A E A E A E A E A E A E A E A E A	V F P D V D Y V N S T F A E T I L D N I T Q S T Q S F N D F T R V V G G E D A K tgttttcctgatgtsgactatgtaaattctactgaaaccattttsgataacatcactcaaagcacccaatcatttaatgacttcactcgggttgttsstggagaagatgccaa 7210 7210 722G 723O 724O 725O 726O 727O 728O 729O 730O 731O	P G 'S E P W Q accassicaticcettsocasstactitatactsatssists tsas consistents of the past of	TGT GAGAD GTATTTA GGCA GCTT CAGCACTAACC AATGTGAGAAGGCCTC CAGAGATGAGCA GTTGGTGAAGAGGGCTCAAAGAGGAGAAGAATTTGGCAT 7450 746G 7470 7230 7550 7560 7560 7560	TAAGSADACAGCATA 6CAG SATTCCAGACAGGCCATTGAACATGAAGGTCTGGAAGAAGGTCGCAGGTACTCAGGCCACTACTTCAGCTTCAGCCTTGCAAAAACT 7573 7580 7590 7590 7600 7610 7620 7630 7640 7650 7650	GGT GAGASTTG GAAA GTCTTTA GGAAAAA TTGGATTATTTAA AA AGGGGGTAA AGGAACTCAAGGAGGAGGGGAGG	AGA GTCTTGATCTAC CACTATAGTTCTCGTGGAACTGAACGGAGATTACT 7870 7880 7890 7900 7910 7920 7920	TAACCG4-aTTTGA-1AC-CCTGG-CAACACG-CGAACCCCACCTCTAATTAAAAAAATACAAAATTAGCTAGGTGTGATGACTCCCACCTGTGCTCCCAGCTATTCAGGAGGCTGAGG 7910 7940 7050 7060 8030 8040	E

0726	9230	9220	9200 9210 9220 9230 9240	9209	9190	9180	9170	9160	9150	9140	7130
9120 TCATAAATG	9110 TCTTATAAC1	9100 6016-ACCGA	9090 Gacaatatiti	0906 444	0406	0906	0506	0706	0206	902Ú	, 010
0006	0668	0868	8970	8960	8950	8940	8930	8920	8910	0068	04 8 F
8380	8870	- I	8850	0788	8330	8820	8310	0.	8760	3 7 8 O	8 7 70
12	5 1	8740	3TAGTAATTTG 8730	AGTTTAGTGTAACTTAAACT-CCTAGTTGCCACAAGTCATGATTTAGTAGTAATTTCATGGA 3689 3590 8730 8710 8720 8730 87	7 TGCCACAAG 8710	ACT-CCT4G 3700	3590 3590		AGCACTTATAA E670	CTATT 60 1	TACTGCAGAAATTECTATTUGATAGCAGTTATAAT SSSSO BOOC
CACTCTAT 8640	36464TAGAA 8630	8620 8620	;TGTATTGGAG 8610	ACCTGCCAAATTTTAAGCACATAACAAC CACATGTGG-TAGTAACTACTGTATTGGAGAGTGCAAGGGGGAGATAGAACACTCTAT 9560 8570 8530 8590 8600 8610 8620 8640	CACATGTGG 8590	ACATAACAA(8530	14 TT TT A A GC. 8 5 7 0	CACCTGCCA4 9560	TTC-TCAATT5 8550	AAAATTCAA1 8540	ATT AJGAGAAATTAAAAATTCAATTG-TCAATTGC 8530 8550
144C14CA 8520	1441TT444T 8510	.TATGTTTA 8500	CACTASCCACA 8490	RABGCCAGSATGGTCCTTATATAAAGTTGTGCTGT-CAATAGGTAACCACTAGCCACATATGTTTAAATTTAACTAGC 3440 8450 8450 8510 8520	1 GTGCTGT-C. 8470	1147846T1 8460	8450	G	CTCCCCAGGCA: 8430	1766CCCCC 8420	AATAACCCACCTCTTTGGCCCCCCTCCCCAGGCAG
16A5AAGA 3400	17CCGTCTG 6390	GCCTAAGAGA 8380	TACAGCTGAA 8370	a a ca a c ca a 6 c ta catitica da a toticicatica a como como como como como como como	A ATCTGGGTT(8350	77GTAGCAA# 8340	6-6CTACAT' 8330	TACCAACCAA 9320	SATTTTCATT. 8310	ACCTATGCAC 830C	ACATATGTAGAATTACCTATGCACATTTTTCATTT 8310 8310
AACAACCT B230	1767CTGCA 8270	TCTCCAAGCG 8260	.TGC AG &CATT 8250	TCABABCACCAAATAATGCACTTGTAC CTASTCCTTCCGGGTGCTCTGCAGACATTTCTCCAASCGTAGTCTGCAAACAACCT 3200 5210 8270 8280 8250 8250 8250	.CTASTCCTT(8230	SCACTTGTAC 9220	CCAAATAAT(5210		. A G A A G C C G G G (8 1 9 0	A ACTTGTTAC 818Ú	SCC CaaaggaaatsaaCTTGTTACAGAAGCCGGGG 8170 8180 8190
								.d.,	·	-	
4085 8090 8100 8110 8120 8130 8140 8150 8160										-	

)TTCACA 9350	9480	AGA A CA G 9600	9720	676A6CC 984D	AGTTT		10080	10200	10320	
STTTACTCTT0 9350	FTCTCTCTAA 9470	TGTGGGACAT 9590	AACATAGTGA 9710	GGAGCTGGCA 9830	-1G1-11G1G 9950		10070	10190	STGAAAGACC 10310	FIG. 7i
CATCCATATO 9340	CATGTATCTAT 9460	rcc1166611 9580	CATCCTGGCT 9700	0286 9820	AAAAGGTGT 9940		10050	10180	.CAGAATAGTT 10300	
.TTCCTTCACT 9330	ITTTAATATCC 9450	CAATTATCT1 9570	16atcaaac) 9690	4TGGCGTGAA0 9810	88888888888888888888888888888888888888		10050	10170	AGCTACTTTC 10290	
CCATAACCCC 9320	CTTCAATGTA 9440	.TACTTTATGA 9560	1 A G A T T C A G G A 9 6 8 0	166C 466AGA1	246TTTAAA. 9920		10040	10160	CAAGAAGCTG 10289	
CCCTTACCTA 9310	GTTATTATGC 9430	CAAAGATCC 9550	74444GTCC4 9670	3668666764 9790	MAAAAAGTCC 9910		10030	10150	TTTCCAAAGG	
AGSACTGCT T 9300	TGCAAGCACT 9420	ATGGTTACA 9 9540	CTTCCGTTT1 9660	CCAGCTACA(9780	88888888888888888888888888888888888888		10020	10140	SCACTTTATC 19260	
TTAATTGCACCCTATGAGSACTGCTTCCCTTACCTACCCTTACCCCTTCCTTCACTCCATATCTTTACTCTTCT	AACAATCCCTCTTAAGTGCAAGCAC 16TTATTATGCCTTCAATGTATTTAATATCCATGTATCTATTCTCTCTAATTTTGTC 00 9410 9420 9430 9440 9450 9460	AACTTCCATGGATAACATGGTTACA ACAAAGATCCTACTTTATGACAATTATCTTCCTTGGGTTTGTGGGCATAGAACAG 20 9530 9540 9550 9560 9570 9580 9590	TTAGCTAAGAAGATAACTTCCGTTTTTAAAAGTCCAAGATTCAGGAGATCAAAACCCTCGGCTAACATAGTGAAACCCCG 40 9050 9660 9670 9680 9690 9700 9720	GCCTATAGTC 9770	GAGACTCCAAAAAAAAAAAAAAAAAAAGTCCAAGTTTAAAAAAAAAA		10010	10130	CCCTATICAACCACATGAACAGATTACTGATGACAGATTCAAAGCACATTATCTTTCCAAAGGCAAGAAGCTGAGCTACTTTCCAGAATAGTTSTGAAAGACCCTGTCAT 10210 10220 10230 10240 10250 10250 10250 10260 10270 10280 10290 10300 10310 10320	
CTACTTAATT 9289	C TGG A A C A A T 9400	GTCCAACTTC 9520	TATATTAGCT 9640	.G.S.TG.C.A.G.G.C 97.60	.G.A.GC.G.A.G.A.C.1 9830		1 0000	10120	1024 D	
1CCTCTSTTT 9279	-GAACCITIC 9390	TTCATTATGT 9510	CCC 463 A GAA 9630	AGCCC5GCGT 9750	.CT655CGaCA	•	0000	10110	IGAACAGATT! 10230	
CCTTTTTACC 9260	ACCTTCTTTA 9380	161 A 1 1 1 1 C A 9 5 0 G	GATCCAAGAA 9326	ACAAAAATT 974U	GCACTCCAGC 9300		0866	10100	FTCAACCACA1 10220	÷
G-ACACTGTATGTTCCTTTTTACCTCCTCTSTTTCTAC	ACTCTGTAATATTGACCTTCTTTA-GAACCTTTCCTGG 9370 9386 9399	ATTITGIGITC TCATGIATITTCATTCATTATGIGTCC 94.40 9500 9510	TGC TCAGAGTAGGGGATCCAAGAACCCAGGAGAATATA 9512 9526 9630	TCTCTTCCAAAAAATACAAAAAATTAGCCCGGCGTGGTGGCAGGCCCTATAGTCCCAGCTACACGGGAGGCTGAGGCAGGC	GAS ATCCCGCCACTGCACTCCAGCCTGGGCGACAGAGC 9350 9350 9300		0206	10090	CCC TA1 10210	
ù-4 C£	ACT C1	ATTT	166 11	1616	GA GA		i			

10440	10560	;TTTCAAAA 10680	10800	17GTGGACT 10920	11040 11040	3CAGGCTAA 11160	11280	11400	SGCTCAGA 11520
AGACAAAATU 10430	CTGCAGCA4/ 10550	GAAAAGAGT(10670	AACATAGAGI 10790	ACAACTTCT 10910	. G G G G C A A A '	GAGATCAGA(11150	111CACCTC/	.TCCTAAAGC	.ccccccaT
AGTCATTTTT 10420	GCAGATCAGA 10540	GTGTCACCTA 10660	CACAAAGAAG 10780	AAAGGGGAAG 10900	666CAGCATT 11020	CACCTGGCCA 11140	ATTCCAGGGC 11260	GTTTGCTACC 11380	CCCTCCAGGC 11500
1616TATAAA 10410	10530	; TGT T T A G C A 10650	SAGGAATCT 10770	3CCACAAGGG 10890	11010	36CATTTGTT 11130	11250	CCCATCCT	ACTTCAGTTT 11490
GATAAATTGA 10400	GGAAATACT4 10520	1 A T G A A A T A C 1 0 6 4 0	.TCACTTGGTC 10760	TTTGCAAGAC 10880	11000	CAGGGAG-G(CACGGTCAT1 11240	.ACGGCGCTG- 11360	.GAAGGASST/
AAAATATGA 10390	TCT GCA GG SA 10510	TAGCTTTGAA 10030	AATAAAGTGA 10750	AGGTGAGCTG 10870	CAACAACCAC 10990	TAGCCCCATT 11110	CGTCTGCGGT 11230	TGGTTCCGGA 11350	GAGGCCAATG 11470
10330	ICATGGTTAA 10500	11141414 10629	24636T666A 10740	34446GTGT 10860	CTTTCCAAC 10980	11100	SCACGTATCC 11220	34GCCCCCTA 11340	364ACGGCGT 11460
AGTICCTTATSAAT3GTTACTGGTTTTCAAAATATGAGATAAATTGAGTGTATAAAAGTCATTTTTAGACAAAATGAAACAGGA 10369 10370 10390 10390 10390 10400	TGGGCCAGCTCCACCATGTCATGGTTAATCTGCAGGAAATACTAGATTTGATTGCAGATCAGACTGCAGCAAACCTGCTGT 19460 10490 10500 10510 10520 10550 10550 10560	GGCTTCAGTGGTGAAACATTATATATCTAGCTTTGAATATGAATACTGTTTAGCAGTGTCACCTAGAAAAGAGTGTTTCAAAA 10600 10610 10620 10630 10640 10650 10660 1068	TTATCTTTCAAATTTAGCCAGSGTGGGAAATAAAGTGATCACTTGGTGAAGAAATCTCACAAAGAAGAACATAGAGAGTTCACTT 10720 10730 10740 10750 10760 10770 10780 10790 10800	AATGGTTAGTCTGTTAAAGAAAAGGTGTAGGTGAGCTGTTTGCAAGACCACACAGGGÄAAGGGGAAGACAACTTCTTTGTGGACT 10340 10350 10860 10870 10890 10890 10900 10900 10900 10920	GACCTCCATTAAGAAA 3CCCTTTCCAACCAACCACTGGTTGGTTACACAGGTTGGGCAGCATTGGGAGCAAATGTTGATTG 10960 10970 10980 10990 11000 11010 11020 11030 11040	PGTTCTGTCACTSGGGACAGCGGGTAGATAGCCCCATTCAGGGAG-GGGCATTTGTTCACCTGGCCAGAGATCAGAGCAGGCTAA 11090 11090 11100 11110 11120 11130 11140 11150 11160	CAGAGCCATGITCICCIAGCACGIATCCCGICIGGGICACGGICATITCTIACCITATICCAGGGCITICACCTCAGCTIGCCA 11200 11210 11220 11230 11240 11250 11280	11330	3TTCCGGACA() 11450
CAGTTCCTT/ 10360	4 TGGCCAG(10600 TCAG	10720	A & A T G S T T A (.T GACC TCC A'	11090	11200 11200	C GCGATGGT	16 GTTAGAAG(
CACCTCCATC 10350	TCATTTGTGG 10470	6CAACASCTG 10590	4GTTGTTTC1 10710	AACAAACTAG 10830	aagacgattc 10950	ACTTABAGAG 11070	TT GAGACCC 1 11190	CGC-CTTGT1	AGTCGCGCA4
TUCTCCACAC 10340	GAATCTCTCC 10460	6 26 2 2 4 6 C 2 4 10 5 3 0	TTCTCTTCAS 10700	G AA CA GATTG 10 b 2 c	GCAAGCAGGC 1094 O	6644T16TT9 11050	C-TG TC CASC T 11180	5 CA AC GCA GC 11 300	CTTCTAA
ACTICISCATIGITILCICCACACCACCICCATCC 10330 10345 10355	AAT 6AAAGAAA CCAC AATCTCTCCTCATTTGTGGA 10450 10450 1046	GACTAAGGCATCAAGAGAAAGCAAGCAACAGCTGG 10570 10530	TGC TGATGCAACCITTCTCITCAGAGTTGTTTCTT 10690 10700 10710	TCATCTGGAGTAATGAACAGATTGAACAACTAGA 10810 10820 10830	TAAGSGTSAAASTTGCAAGCAGGCAAGACGATTCT 10930 10940	AAC AAAT GTTT GTCG GAATTGTT GACTTAAAGAGC 11050 11060 11070	66-aCT-CT66aTCCTGTCCASCTTTGAGACCCTA	GGC TSSAGCCA AGGS CAACGCAGCCCCC CTTGTTC GCGATGCTTCCCAGGAGCCCCCTATGGTTCCGGACGGCGCTG-CCCATCCTGTTTGCTACCTCCTAAAGCCAAAGGC 11230 11300 11310 11320 11330 11330 11340 11350 11360 11360 11370 11380 11380 11400	TGGCGGG-C-GG-CCTICTAAAGTCGCGCAAGGTTAGAAGGTTCCGGACAGGAACGGCGTGAGGCAATGGAAGGAGGAGTACTTCAGTTCCCTCCAGGCCCGCGATGGGCTCAGA 11410 11420 11430 11440 11450 11440 11460 11470 11480 11480 11500 11510 11520
. AC T T	AA T G	GACT	Tuc T	TC A T(TAAG	AACE	. 55 - A	66.0	1660

FIG. 7k

4 D	⊢ 0	
1164	1176	12
AGTTTTCT1 11630	SCAACGCCT(11750	CTGCCAGCT
AACAGAATG 11620	AAGGGAAC1 11740	GCCCATGAC(
ACTACACAGG 11610	TTAGTGAAAA 11730	TGGATCCCAT
36TTAAATGC 11600	CCTTTGAATA 11720	1441116611
5a a gCT a g a g (GAAATGAAAA 11710	TCT A A B G C T T
11580	11700	111166611
ACTTCAGSAG 11570	11690	.c a a c a c c 6 C 4
с тбаабава т 11560	TCTTGAATTG 11680	TTTAAAGAAG
GCTCCTTSAGAAÖTC GGGAAGGAAGGACTCTCTGAAGAATACTTCAGSAGTAGAAGAGGAAGCTAGAGGGGTTAAATGCACTACACAGAACAGAAATGAGTTTTCTTAGAGTTA 11530 11540 11550 11550 11560 11570 11590 11590 11600 11610 11620 11630 11640	GTATATGTCTASASGTGTAGTAAACTAAAACTACCACTACCGCCACGTAGGGAAGAAATGAAAACCTTTGAATATTAGTGAAAAAAGGGAAACTSCAACGCCTGTATTACT 11655 11663 11670 11750 11760 11760	AS A TAGC TITC A TCA ACUG CTC A A A A CC SAC A GA TITA A A GA A SC A A CACC GCATTITG SCTTT COTA A TOTA A TITGG TTT GG A TCC CATGC CCATGA CCTGC CAGCT G
. 666 A A A G G A 115 & O	GTGTAGTAAA 11003	a aC u G CTC A A
.TTSAGAAČT(11530	1161CTA 545	16C T T T C & T C.
32139	GTATA	26 A T 3

BNSDOCID: <GB___2125409A_1.>

FIG. 8(a)

			110. 0(1)
→ 1	0.000	ECOR1	GAATTC
30	0.002	HINF1	GAATC
33	0.003	MB 011	TCTTC
46	0.004	ALU1	AGCT
48	0.004	DDE1	CTGAG
50	0.004	MNL1	GAGG
8 9	0.007	MNL1	CCTC
94	0.008	MST1	TGCGCA
95	0.008	HHA1	GC G C
112	0.009	MB 01	GATC
120	0.010	98V1	GC A G C
120	0.010	FNU4H1	GCAGC
123	0.010	88 V1	GC A GC
123	0.010	FNU4H1	GCAGC
134	0.011	DDE1	CTGAG
148	0.012	HPH1	GGTGA
173	0.014	MNL1	GAGG
188	0.016	DDE1	CTTAG
204	0.017	HINF1	GAATC
247	0.021	SPH1	GCATGC
265	0.022	ALU1	AGCT
266	0.022	BBV1	GCTGC
266	0.022	FNU4H1	GCTGC
305	0.026	XMN1	GAACACTTTC
376	0.032	ALU1	AGCT
417	0.035	MNL1	GA G G
425	0.036	STU1	AGGCCT
426	0.030	HAE111	GGCC
465	0.039	RSA1	GTAC
488	0.041	DDE1	CTTAG
517	0.043	ALU1	AGCT
523	0.044	ALU1	AGCT
559	.0.047	MNL1	ССТС
578	0.049	RSA1	GTAC
590	0.050	DDE1	CTAAG
621	0.052	ALU1	AG C T
652	0.055	HINF1	GATTC
→ 732	0.062	HIND111	AAGCTT
733	0.062	ALU1	AGCT
781	0.066	MB 011	GAAGA
788	0.066	MNL1	GAGG
816	0.069	MNL1	GAGG

FIG. 8(b)

818	0.069	FOK1	GGATG
898	0.076	MNL1	CCTC
898	0.076	MST11	CCTCAGG
899	0.076	DDE1	CTCAG
913	0.077	DDE1	CTGAG
929	0.078	HPH1	GGTGA
976	0.082	TAQ1	TCGA
1027	0.036	RSA1	GTAC
1032	0.087	MNL1	GAGG
1054	0.039	MNL1	CCTC
1072	0.090	HIND111	AAGCTT
1073	0.090	ALU1	AGCT
1073	0.092	BBV1	GCAGC
1099	0.092	FNU4H1	GCAGC
1101	0.092	ALU1	AGCT
	0.096	MNL1	GAGG
1138	0.096	HINC11	GTTGAC
1145 1150	0.097	FOK1	CATCC
	0.097	ALU1	AGCT
1161 1167	0.098	HPH1	TCACC
	0.100	HPH1	GGTGA
1193	0.101	ALU1	AGCT
1198		DDE1	CTGAG
1200	0.101	MB011	GAAGA
1204		MNL1	GAGG
1226	0.103	DDE1	CTGAG
1284	0.108	MNL1	GAGG
1286	0.108	RS41	GTAC
1323	0.111	88V1	GCTGC
1365	0.115	FNU4H1	GCTGC
1365	0.115	XBA1	TCTAGA
1370	0.115	DDE1	CTAAG
1424	0.120	ALU1	AGCT
1427	0.120	RSA1	GTAC
1449	0.122		AGCT
1603	0.135	ALU1	GTATAC
1626	0.137	ACC1	GTTAAC
1633	0.137	HINC11	GTTAAC
1633	0.137	HPA1	GAGG
1670	0.141	MNL1	GGCC
1672	0.141	HAE111	GGATG
1685	0.142	FOK1	GATTC
1759	0.148	HINF1	
1766	0.149	MNL1	GAGG
1841	0.155	SAU961	GGGCC
1842	0.155	HAE111	GGCC

1855	0.156	DDE1	FIG. 8(c) CTTAG TCTTC GGACC GGACC CCTC CTCAG
1884	0.159	MBO11	
1901	0.160	AVA11	
1901	0.160	SAU961	
1939	0.163	MNL1	
1940	0.163	DDE1	
1945 1965 2081 2097 2111 2112 2111 2114 2115 2116 2116 2116 2116 2117 2116 2117 2117	0.164 0.165 0.165 0.177 0.178 0.178 0.178 0.179 0.181 0.181 0.181 0.182 0.182 0.182 0.183 0.183 0.187 0.187 0.187 0.192 0.192 0.193 0.193 0.193 0.194 0.198 0.208 0.208 0.208 0.211 0.212 0.213	ALE111 SASA1 HAUSA1 RESA1 HAUSA1	AGCT GGCC GGCCC GTAC GTAC GACGC AGCT CTCAG GATC CCTC CC

			FIG. 8(d)
2658 2678 2772 2777 2815 2988 2988 2988 2988 2988 3038 3093 3146 3193 3216 3226 3233 3233 3216 3233 3233 323	0.224 0.225 0.230 0.230 0.233 0.236 0.237 0.250 0.251 0.251 0.255 0.255 0.255 0.260 0.260 0.260 0.262 0.264 0.267 0.269 0.271 0.271 0.271 0.272 0.275 0.281 0.287	RSA1 SFNA1 HINC11 HINC11 HINF1 HGA1 DDE1 HINF1 AVU961 MNNF1 ALU1 NDE1 MNDE1 MBO11 RSA1 TRSA1 MBO11 RSA1	GTAC GCATC GAGTC GAGTC GATTC GACTC GATTC GATTC GATTC GATTC GATTC GATTC GAGT GATTC GAGT GATTC GAGG GAAC GTAC GT
3413 3415 3457 3462 3489 3522 3585 3624 3625 3638 3689 3792	0.287 0.288 0.291 0.292 0.294 0.297 0.302 0.305 0.305 0.311	MB01 HPH1 DDE1 HINF1 TAQ1 ECOR5 RSA1 BGL11 MB01 MB01 MB01 HPH1 ALU1	GATC TCACC CTAAG GACTC TCGA GATATC GTAC AGATCT GATC GATC

			FIG. 8(e)
38 4 7	0.324	RSA1	GTAC
3905	0.329	RSA1	GTAC
3970	0.334	BSTN1	CCAGG
3970	0.334	SCRF1	CCAGG
3979	0.335	BSTE11	GGTAACC
4016	0.338	MNL 1	GAGG
4022	0.339	SFNA1	GCATC
4025	0.339	MB011	TCTTC
4368	0.368	HINF1	GAGTC
4384	0.369	RSA1	GTAC
4410	0.371	SFNA1	GATGC
4469	0.376	SFNA1	GATGC
4520	0.381	RSA1	GTAC
45 2 3	0.381	DDE1	CTGAG
4525	0.381	MNL1	GAGG
45 2 9	0.381	ECOR5	GATATC
4533	0.382	TAG1	TCGA
4658	0.392	HINF1	GAATC
4695	0.395	ALU1	AGCT
4719	0.397	XBA1	TCTAGA
4727	0.398	SFNA1	GCATC
4769	0.402	ECOR1	GAATTC
4769	0.402	XMN1	GAATTCTTTC
4778	0.402	DDE1	CTGAG
478.0	0.403	HINF1	GAGTC
4848	0.408	NDE1	CATATG
4961	0.418	HINF1	GATTC
4988	0.420	DDE1	CTGAG
5020	0.423	ALU1	AGCT
5022	0.423	DDE1	CTGAG
5049	0.425	HINF1	GATTC
5053	0.426	HP411	CCGG
5085	0.428	BCL1	TGATCA
.5086	0.428	MB01	GATC
→ 5157	0.434	PVU11	CAGCTG
51 5 8	0.434	ALU1	AGCT
5225	0.440	ACC1	GTAGAC
5258	0.443	PST1	CTGCAG
5285	0.445	MNL1	GAGG
5339	0.450	ECOR5	GATATC
5355	0.451	RSA1	GTAC
5367	0.452	HGIA1	GTGCAC
5394	0.454	RSA1	GTAC
5402	0.455	DDE1	CTCAG
5414	0.456	BSTN1	CCAGG

FIG. 8(f)

			110. 0(1)
5414	0.456	SCRF1	CCAGG
5421		MB011	GAAGA
5451		MB01	GATC
5455		ALU1	AGCT
		•	
E . 24	0.462	FNU4H1	GC G G C
5481 5490		MNL1	GAGG
5560		ALU1	AGCT
5562		DDE 1	CTGAG
5627		XMN1	GAAAGTATTC
5653		FOK1	GGATG
5657	-	HINF1	GAGTC
5672		PST1	CTGCAG
5674		BBV1	GC A G C
5674	0.478	FNU4H1	GCAGC
5754	0.485	BSTN1	CCTGG
5754		SCRF1	CCTGG
5761		SAU961	GGGCC
5762		HAE111	GGCC
5764		BSTN1	CCAGG
5764		SCRF1	CCASG CCTC
5779		MNL1	GATATC
5813		ECOR5 Hae111	GGCC
5621		88V1	GCTGC
5844 5844		FNU4H1	GCTGC
5845		PST1	CTGCAG
5863		BAL1	TGGCCA
5864		HAE111	GGCC
5875		SAU961	GGGCC
5876		HAE111	GGCC
5886		9 A L 1	TGGCCA
5887		HAE111	GGCC
5898	B 0.497	MNL1	GA G G
5899	9 0.497	STU1	AGGCCT
5900	0.497	HAE111	GGCC
5922	2 0.499	ALU1	AGCT
5952		MBO11	GAAGA
595		HINF1	GAATC
5961		DDE1	CTAAG
597	1 0 . 503	SAU961	GGGCC

			FIG. 8(g)
5972 5987 5994 6000 6021 60267 6121 6139 6177 6211 6214 6233 6249 6275 6379 6379 6380 6381 6558	0.503 0.504 0.505 0.505 0.507 0.507 0.508 0.515 0.517 0.520 0.523 0.523 0.523 0.525 0.526 0.526 0.528 0.528 0.528 0.537 0.537 0.537 0.537 0.552	HAE111 MB011 BSTN1 SCRF1 MB011 ALU1 ACC1 MNL1 ALU1 MB011 MNL1 DDE1 ALU1 HAE111 HIND111 AVA11 SAU961 RSA1 MB011 BBV1 FNU4H1 PVU11 ALU1 AVA11	GGCC TCTTC CCTGG CCTGG TCTTC AGCT GTCTAC GAGG AGCT TCTTC CCTC CTTAG AGCT GGCC AAGCT GGCC CTTC GCAGC GCAGC CAGCT GCAGC CAGCT GCAGC CAGCT GGTCC
6558 6561 6561 6564 6629 6639 6674 6677 6683 6684 6722 6722 6767 6793	0.552 0.553 0.553 0.558 0.559 0.562 0.562 0.563 0.563 0.566 0.566	SAU961 BSTN1 SCRF1 HPH1 HINF1 MB01 HINF1 XBA1 STU1 HAE111 BBV1 FNU4H1 SFNA1 FOK1 HINF1	GGTCC CCTGG CCTGG GGTGA GAATC GAATC TCTAGA AGGCCT GGCC GCAGC GCAGC GCAGC GCATC GGATC

FIG. 8(h)

6874	0.579	HINF1	GATTC
6911		ECOR1	GAATTC
6916		HPA11	CCGG
6984		ALU1	AGCT
6991	0.589	HINF1	GACTC
7028	0.592	SAU961	GGGCC
7029	0.592	HAE111	GGCC
7038	0.593	DDE1	CTCAG
7052	0.594	FCK1	GGATG
7056		S AU 961	GGGCC
7057		HAE111	GGCC
7059	0.594	MNL1	CCTC
7124	0.600	MBO11	TCTTC
7155	0.603	M5011	GAAGA
71 5 5	0.603	XMN1	GAAGAGTTTC
7179	0.605	DDE1	CTAAG
7182	0.605	ALU1	AGCT
7185	0.605	HPH1	TCACC
7194	0.606	DDE 1	CTGAG
7196		MNL1	GAGG
7237	0.609	ALU1	AGCT
7293	0.614	A V A 1	CTCGGG
7310	0.616	MB011	GAAGA
7313	0.616	SFNA1	GATGC
7322	0.617	ESTN1	CCAGG
7322	0.617	SCRF1	CCAGG
7343	0.618	RSA1	GTAC
7373	0.621	HGIA1	GAGCTC
7373	0.621	SAC1	GAGCTC
7374	0.621	ALU1	AGC T
7376	0.621	DDE1	CTCAG
→ 7378	0.621	PVU11	CAGCTG
7379	0.621	ALU1	AGCT
7394	0.623	HAE111	36CC
7396	0.623	BSTN1	CCAGG CCAGG
7396	0.623	SCRF1	
	0.624	ODE1	CTGAG
	0.624	MNL1	GAGG
	0.626	FOK1	GGATG
7485	0.630	STU1	AGGCCT
7480	0.630	HAE111	GGCC CCTC
7488	0.631	MNL1	
7507	0.632	HPH1	GGTGA
7516	0.633	MNL1	GAGG
7529	0.634	ALU1	AGCT
7547	0.636	MRO11	GAAGA

RIC	21	÷)
FIG.	01	1	,

7580	0.638	HINF1	GATTC
7599	0.640	HINC11	GTCAAC
7619	0.642	MB011	GAAGA
7634	0.643	RSA1	GTAC
	0.643	DDE1	CTCAG
7637		ALU1	AGCT
7659	0.645 0.647	HPH1	GGTGA
7681		DDE1	CTAAG
7705	0.649	HINF1	GACTC
7745	0.652	MNL1	GAGG
7753	0.653	HINF1	GAGTC
7802	0.657	MBO1	GATC
7809	0.658	SSTN1	CCTGG
794G	0.669	SCRF1	CCTGG
7940	0.669	MNL1	CCTC
7963	0.671	ALU1	AGCT
7989	0.673		GACTC
8002	0.674	HINF1	GTGCTC
3013	0.675	HGIA1	AGCT
8021	0.675	ALU1	GAGG
8031	0.676	4NL1	CTGAG
8035	0.677	DDE1	GAGG
8037	0.677	MNL1	
8046	0.678	HINF1	GAATC
8049	0.678	HPH1	TCACC
8053	0.678	9 D E 1	CTGAG
3058	0.679	BSTN1	CCTGG
8058	0.679	SCRF1	CCTGG
8067	0.579	TAQ1	TCGA
8069	0.680	MNL1	GAGG
.8072	0.680	BBV1	GCTGC
8072	0.680	FNU4H1	GCTGC
3073	0.680	PST1	CTGCAG
. 8086	0.681	BCL1	TGATCA
8087	0.681	MBO1	GATC
8109	0.683	DDE1	CTGAG
8160	0.687	HAE111	GGCC
816C	0.687	SAU961	GGCCC
8190	0.690	HPA11	CCGG

				FIG. 8(j)
8	190	0.690	NCI1	ccese
8	190	0.690	SCRF1	CCGGG
	3220	0.692	RSA1	GTAC
	1233	0.693	AVA1	CCCGGG
ê	3233	0.693	NCI1	cccgg
	3233	0.693	SCRF1	CCCGG
	1233	0.693	SMA1	ccceee
Ξ	3234	0.693	HPA11	CCGG
	1234	0.693	NCI1	CCGGG
	3234	0.693	SCRF1	CCGGG
	1238	0.694	HGIA1	GTGCTC
	3243	0.694	PST1	CTGCAG
	282	0.697	NDE1	CATATG
	357	0.704	DDE1	CTTAG
	366	0.705	PVU11	CAGCTG
	367	0.705	ALU1	AGCT CTAAG
	376	0.705	DDE1 HINF1	GATTC
	3382	0.706	MB011	GAAGA
		0.707 0.708	MNL1	CCTC
		0.709	HAE111	GGCC
	3417		SAU961	GGCCC
	3423	0.709	MNL1	CCTC
	, ,	0.107		40. 5
8	3428	0.710	BSTN1	CCAGG
8	428	0.710	SCRF1	CCAGG
8	3440	0.711	BSTN1	CCAGG
8	3440	0.711	SCRF1	CCAGG
	1443	0.711	FOK1	GGATG
	3447	0.711	AVATT	GGTCC
	3447	0.711	SAU961	GGTCC
	477	0.714	BSTE11	GGTAACC
	1492	0.715	NDE1	CATATG
	643	0.728	PST1	CTGCAG
	221	0.777	MB01	GATC
	263	0.780	MNL1	CCTC
	266	0.780	MNL1	COTC
	294	0.783	MNL1	GAGG
	335	0.786	FOK1	CATCC TCTTC
. 5	350	0.787	MB 0 1 1	16116

FIG. 8(k)

9353	0.788	MB011	TCTTC
9394	0.791	BSTN1	CCTGG
9394	0.791	SCRF1	CCTGG
9400	0.792	MNL1	CCTC
95 5 C	0.804	MB01	GATC
95 7 1	0.806	MB011	' TCTTC
9600	0.808	HGIA1	GTGCTC
9603	0.309	DDE1	CTC AG
→ 9614	0.810	SAMHT	GGATCC
9615	0.310	MB01	GATC
9626	0.811	BSTN1	CCAGG
9626	0.811	SCRF1	CCAGG
9641	0.812	ALU1	AGCT
9643	0.812	DDE1	CTAAG
9647	0.812	MB011	GAAGA
9676	0.815	HINF1	GATTC
9685	0.816	MB01	GATC
9694	0.816	FOK 1	CATCC
9697	0.817	BSTN1	CCTGG
9697	0.317	SCRF1	CCTGG
9723	0.819	MB011	TCTTC
9747	0.821	NCI1	cccee
9747	0.821	SCRF1	CCCGG
9748	0.821	HPA11	CCGG
9762	0.822	HAE11	GGCGCC
9762	0.822	NAR1	GCGC
9763	0.822	HHA1	AGCT
9777	0.823	ALU1	GAGG
9787	0.824	MNL1	CTGAG
9791	0.825	DDE1 MNL1	GAGG
9793	0.825	HPA11	CCGG
981 4	0.826	NCI1	CCGGG
981 4	0.826 0.826	SCRF1	CCGGG
981 4 981 9	0.827	MNL1	GAGG
9826	0.828	ALU1	AGCT
9843	0.829	MBO1	GATC
9864	0.831	BSTN1	CCTGG
9864	0.831	SCRF1	CCTGG
9881	0.832	HINF1	GACTC
10246	0.863	HINF1	GATTC
10279	0.866	ALU1	AGCT
10281	0.866	0051	CTGAG
10284	0.866	ALU1	AGCT
10310	0 868	TTH1111 '	GACCCTGT

FIG. 8(L)

10336	0.870	MNL 1	CCTC
10347	0.371	MNL1	CCTC
10351	0.872	FOK1	CATCC
10455	0.880	HIMF1	GAATC
10463	0.881	MNL1	· CCTC
10473	0.882	FOK1	GGATG
10477	0.882	SAU961	GGGCC
10478	0.382	HAE111	GGCC
10482	0.883	ALU1	AGCT
10505	0.885	P S T 1	CTGCAG
10512	0.885	MNL1	GAGG
10536	0.387	M301	GATC
10543	0.888	PST1	CTGCAG
10545	0.338	B3V1	GCAGC
10545	0.888	FNU4H1	GCAGC
10563	0.890	0051	CTAAG
10568	0.890	SFNA1	GCATC
10589	0.392	PVU11	CAGCTG
10590	0.892	ALU1	AGCT
10605	0.393	HPH1	GGTGA
10625	0.595	ALU1	AGCT
10656	0.897	HPH1	TCACC
10685	0.900	SFNA1	GATGC
10693	0.901	MB 011	TCTTC
10733	0.904	BSTN1	CCAGG
10733	0.904	SCRF1	CCAGG
10751	0.905	BCL1	TGATCA
10752	0.905	M = 01	GATC
10760	0.906	HPH1	GGTGA
10763	0.906	MBO11	GAAGA
10779	0.908	M5011	GAAGA
10865	0.915	HPH1	GGTGA
10869	0.915	ALU1	AGCT
10899	0.918	MB011	GAAGA GGTGA
10925	0.920	4PH1	GATTC
10950	0.922	HINF1	CCTC
10958	0.923	MNL1 83V1	GCAGC
11015	0.928	22 A 1	GC M G Ç

FIG.	8(m)
------	------

44545	0 030	FNU4H1	GCAGC
11015	0.928 0.932	HINC11	GTTGAC
11061		ALU1	AGCT
11073	0.933	FNU4H1	GCGGC
11095	0.934	HPH1	TCACC
11132	0.938	BSTN1	CCTGG
11135	0.938		CCTGG
11135	0.938	SCRF1	TGGCCA
11137	0.938	9AL1	GGCC
11138	0.938	HAE111	GATC
11145	0.939	MB01	CTAAG
11157	0.940	DDE1	GGATCC
11170	0.941	BAMH1	GATC
11171	0.941	MBO1	AGCT
11181	0.942	ALU1	
11256	0.948	BSTN1	CCAGG
11256	0.948	SCRF1	CCAGG
11265	0.949	нрн1	TCACC
11268	0.949	MNL1	CCTC
11269	0.949	DDE1	CTCAG
11272	0.949	ALU1	AGCT
11278	0.950	BSTN1	CCAGG
11278	0.950	SCRF1	CCAGG
11300	0.952	3 B V 1	GC A G C
		\$	
1130G	0.952	FNU4H1	GCAGC
11303	0.952	FNU4H1	GCCGC
11314	0.953	NRU1	TCGCGA
11315	0.953	FNUD11	CGCG
11324	0.954	ALU1	AGCT
11336	0.954	BSTN1	CCAGG
11330	0.954	SCRF1	CCAGG
11349	0.956	HPA11	CCGG
11356	0.956	HAE11	GGCGCT
11357	0.956	HHA1	GCGC
11367	0.957	FOK1	CATCC
11381	0.958	MNL1	CCTC
11428	0.962	FNUD11	CGCG
11429	0.963	HHA1	GCGC
11447	0.964	HPA11	CCGG
11447	0.965	MNL1	GAGG
11404	U • 70 J		

FIG. 8(n)

11466	0.966	HAE111	วัวออ
11478	0.967	MNL1	GAGG
11481	0.967	RSA1	GTAC
11494	0.968	MNL1	CCTC
11497	0.768	BSTN1	CCAGG
11497	0.968	SCRF1	CCAGG
1150C	0.768	HAE111	GGCC
11500	0.968	SAU961	GGCCC
11504	0.969	FNUD11	CGCG
11505	0.969	HHA1	GCGC
11506	0.969	FNU011	CGCG
11515	0.970	DDE1	CTCAG
11519	0.970	HGIA1	GAGCTC
11519	0.970	SAC1	GAGCTC
11520	0.970	ALU1	AGCT
11533	0.971	AVA1	CTCGGG
11557	0.973	MB011	GAAGA
11560	0.974	XMN1	GAAATACTTC
11581	0.975	MNL1	GAGG
11586	0.976	ALU1	AGCT
11591	0.976	MNL1	GAGG
11631	0.980	DDE1	CTTAG
11648	0.981	XBA1	TCTAGA
11552	0.981	MNL1	GAGG
11701	0.985	MB011	GAAGA
11765	0.991	ALU1	AGCT
11778	0.992	ALU1	AGC T
 ▶11828	0.996	HIND111	AAGCTT
11829	0.996	ALU1	AGCT
11845	0.998	BAMH1	GGATCC
11846	0.998	MBO1	GATC
 11868	0.999	PVU11	CAGCTG
11869	1.000	ALU1	AGCT

V F P D V TGTTTTTCTGATG 590 600 CTI 66C A CETTETITICA A TEKNET OF TATE A A TEKNET OF TATES A A TETTET OF TETTE OF TATES OF TETTE OF TATES OF TETTE OF TATES OF THE TATE ¥ S F N D F T R V V G G E D A K P G G F F CATTAATGACTCCAGGTCAATTCC CATTAATGACAGGTCAATTCC 670 690 700 710 720 ₹. IGGACTATGTAAATTLTAGTAAACTATTTGGATAACATCAGTCAAAGCACCCAATC **8** -≥ 4 7 650

æ 6 FIG.

(d) 6 (P)

320FMKGRS 1 TVNNMFLLE CENSINKE FT TVNNMFLLE CENSINKE FT TVNNMFLLE CENCENCE CONTRACA CATETIC CENCENTE CONTRACA CATETIC CENCENTE CONTRACA CATETIC CATE M E G G K D S C G D S G G P M V T E V E G T S F L T G I I S M G E E C A M K K (TCCLTGABGGAGGTABGAGAGAGAGAGAGGGACCCCATGTTACTGAAGTGGAAGGGACCAGTTTCTTAACTGGAATTATTAGCTGGGGTGAGAGTGTGCAATGAAAG 1210 1220 1230 1240 1250 1250 1250 1250 1250 1250 R Y S I Y T K V S K Y V N W I K E K T K L T # Gcaartatggaatatataccaagstatcccggtatatgaataataaaaacaaagcicacttaatgaaagatggatticcaaggttaattcattggaattgaaaattaacagg 1230 1340 1350 1350 1360 1370 1380 1390 1400 1410 1420 1430 GCC. TCT CACTA ACTA ATT CCCAT CTTTTG ATTA TATA CATT CTATG ATT GCTTTTT CTTTA CAGGGGGGGAGA ATT CCATATITT ACCTGAGCAAATT GATTA GAT AAAAIGGBBCCBCIBGAGGBATATABTGTTTAGGBBTTACAGTCATTTCTAAGGGCCCAGC CCTTGACAAATTGTGAGGTTAAATTCTCCACTCTGTCCATCAGATACTATGGTTCT 1570 1580 1660 1600 1610 1670 1680 415

FIG. 9(c)

AGA GOTT GA GOTT GO COTAGA GO CATA A GOTT COTAT TO TOTAGA TO COTAGA GO TO COTAGA GO COTAGA GO TO COTAGA GO TO CO TO COTAGA GO COTAGA GO TO COTAGA GO TO COTAGA GO COTA AGTIGICCITTICIG GITTCGTGTTCCATGGAACATTTTGATTATAGTTAATCCTTCTATC TTGAATCTTCTAGAGAGTTGCTGACCAACTGACGATGTTTCCCTTTGTGAATTAAT 2650 2730 2740 2750 2750 2760 AGTIDITITIBI BIDI GIDA IDIDITATATATATATATA IDA TATACA ATATAD DI CATATG TGTG TGTG TGTG TGTG TGTG TGC ACACACACACACACATATAATGGAAGCAATAA GCCATTCTABGAECTIGTATGETTATGSAGGTCTGACTAGGCATGATTTGACGAAGGCAAGATTGGCATATCATTGTAACTAAAAAGCTGACATTTGACCCAGACATATTGTACTCTTTC 2250 2300 2310 2380 2390 2400 2230 2470 2220 2210 2200 2130

AAACIEGIETTCIGGITCAAA

		<u>Bam</u> HI	PvuII	<u>Hind</u> III	
Oligo N3	5'	GATO	CAGCTGA	31	
			•••••		
Oligo N4		3'	GTCGACTTCG	₽A	5'

Fig. 10

Eco RI

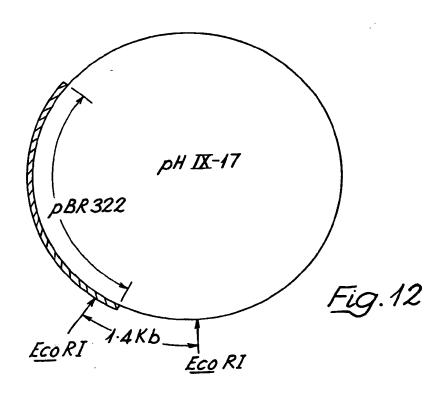
10 20 1 30 140 50

5' GAA TTCTCATGTT TGACAGCTTA TCATCGATAA GCTTCAGCTG GATCCTCTAC

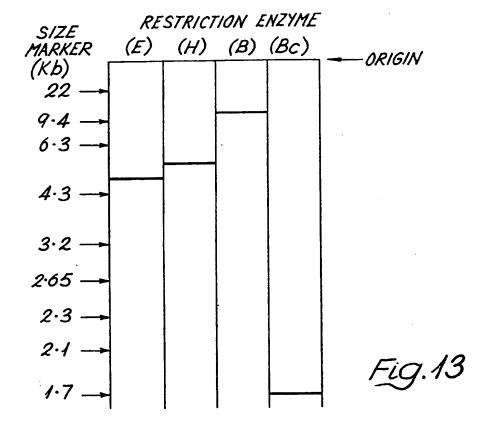
60

GCCGGACGCA 3'

Fig.11



35/35



SPECIFICATION Genetic engineering

BACKGROUND OF THE INVENTION

1. Field of the invention

This invention is in the field of genetic engineering relating to factor IX DNA.

2. Description of prior art

Factor IX (Christmas factor or antihaemophilic factor B) is the zymogen of a serine protease 10 which is required for blood coagulation via the intrinsic pathway of clotting (Jackson & Nemerson, Ann.Rev.Biochem. 49, 765-811, 1980). This factor is synthesised in the liver and requires vitamin K for its biosynthesis (Di Scipio & 15 Davie, Biochem. 18, 899-904, 1979).

Human factor IX has been purified and characterised, but details of the amino acid sequence are fragmentary. It is a single-chain glycoprotein, with a molecular weight of 20 approximately 60,000 (Suomela, Eur.J.Biochem. 71, 145-154, 1976). Like other vitamin Kdependent plasma proteins, human factor IX contains in the amino-terminal region approximately 12 gamma-carboxyglutamic acid 25 residues (Di Scipio & Davie, Biochem. 18, 899---904, 1979)

During the clotting process, and in the presence of Ca++ ions, factor IX is acted upon by activated factor IX (IXa) by the cleavage of two internal 30 peptide bonds, releasing an activation glycopeptide of 10,000 daltons (Di Scipio et al.. J.Clin. Invest. 61, 1528—1538, 1978). The activated factor IX (IXa) is composed of two chains held together by at least one disulphide 35 bond. Factor IXa then participates in the next step in the coagulation cascade by acting on factor X in 100 the presence of activated factor VIII, Ca++ ions, and phospholipids (Lindquist et al., J.Biol.Chem. 253, 1902—1909, 1978).

40 Individuals deficient in factor IX (Christmas disease or haemophilia B) show bleeding symptoms which persist throughout life. Bleeding may occur spontaneously or following injury. This may take place virtually anywhere. Bleeding into 45 the joints is common, and after repeated haemorrhages, may result in permanent and crippling deformities. The condition is a sex-linked disorder affecting males. Its frequency in the population is approximately 1 in 30,000 males.

The current method of diagnosing Christmas disease involves measurement of the titre of factor IX in plasma by a combination of a clotting assay and in immunochemical assay. Treatment of haemorrhage in the disease consists of factor IX 55 replacement by means of intravenous transfusion of human plasma protein concentrates enriched in factor IX. The enrichment of plasma in factor IX is a time-consuming process.

Summary of the invention

After considerable research and experiment, 60 important progress has now been made towards producing artificial human factor IX by

recombinant DNA technology (genetic engineering). Thus, the cloning of DNA sequences 65 which are substantially the same as extensive sequences occurring in the human factor IX genome has been achieved.

The invention arises from the finding that an extensive DNA sequence of the human factor IX 70 genome can be obtained by a clever and laborious combination of chemical synthesis and artificial biosynthesis, starting from elementary nucleotide or dinucleotide "building blocks", as will be described below.

A major feature of the invention comprises 75 recombinant DNA which comprises a cloning vehicle DNA sequence and a sequence foreign thereto (i.e. foreign to the vehicle) which is substantially the same as a sequence occurring in the human factor IX genome. A 11873 nucleotide long part of such a foreign sequence has been identified and a very large part of it has been sequenced by the Maxam-Gilbert sequencing method. A 129 nucleotide length of this sequence is more than sufficient to characterise it unambiguously as coding for a specific protein and a particular such length is regarded herein as useful to characterise the whole sequence inserted in the cloning vehicle as one occurring in the human factor IX genome. Other cloned 90 sequences can then be verified as belonging to the human factor IX genome by determining that part thereof is identical to a region of the firstmentioned sequence, i.e. the sequences have a 95 common identity in an overlapping region.

A further feature of the invention therefore comprises recombinant DNA which comprises a cloning vehicle or vector DNA sequence and a DNA sequence foreign thereto which consists of or includes substantially the following sequence of 129 nucleotides (which should be read in rows of 30 across the page):-

ATGTAACATG TAACATTAAG AATGGCAGAT GCGAGCAGTT TTGTAAAAAT AGTGCTGATA 105 ACAAGGTGGT TTGCTCCTGT ACTGAGGGAT ATCGACTTGC AGAAAACCAG AAGTCCTGTG (1) **AACCAGCAG**

The invention includes particularly recombinant DNA which comprises a cloning vehicle DNA 110 sequence and a sequence foreign to the cloning vehicle, wherein the foreign sequence includes substantially the whole of an exon sequence of the human factor IX genome. The 129-nucleotide sequence described above corresponds 115 substantially to such an exon sequence. Another such exon sequence which independently characterises the human factor IX DNA is the 203nucleotide sequence substantially as follows (again reading in rows of 30 across the page):-

TGCCATTTCC ATGTGGAAGA GTTTCTGTTT CACAAACTTC TAAGCTCACC CGTGCTGAGG CTGTTTTTCC TGATGTGGAC TATGTAAATT CTACTGAAGC TGAAACCATT TTGGATAACA 5 TCACTCAAAG CACCCAATCA TITAATGACT TCACTCGGGT TGTTGGTGGA GAAGATGCCA AACCAGGTCA ATTCCCTTGG CAG

The intron sequences of the human factor IX genome are excised during the transcription 10 process by which mRNA is made in human cells. Only exon sequences are translated into protein. DNA coding for factor IX has been prepared from human mRNA. This cDNA has been partly sequenced and found to contain the same 129-15 and 203-nucleotide sequences set out above.

The invention also includes recombinant DNA which comprises a cloning vehicle sequence and a DNA sequence foreign to the cloning vehicle, wherein the foreign sequence comprises a DNA 20 sequence which is complementary to human factor IX mRNA. Such a recombinant cDNA can be isolated from a library of recombinant cDNA clones derived from human liver mRNA by using an exon of the genomic human factor IX DNA (or 25 part thereof) as a probe to screen this library and thence isolating the resulting clones.

The invention also includes recombinant DNA in which the foreign sequence is any fragment of human factor IX DNA, particularly of length at least 50 and preferably at least 75 nucleotides or base-pairs. It includes such recombinant DNA whether or not part of the 129 or 203-base-pair sequence defined above. It includes especially part or all of the exon sequences of human factor IX 35 genomic DNA. Various short lengths up to about 11 kilobases (11,000 nucleotides or base-pairs) long have been prepared by use of various restriction endonucleases. Methods of isolating recombinant DNA from clones are well known and some are described hereinafter. The DNA of the invention can be single or double stranded form.

The recombinant human factor IX DNA of this invention is useful as a tool of recombinant DNA technology. Thus it is useful as the first stage in 45 the production of artificial human factor IX and in the preparation of probes for diagnostic purposes.

In the production of the artificial human factor IX it is contemplated that appropriate cDNA or genomic clones will be introduced into a suitable 50 expression vector in either mammalian or bacterial 115 produced for examining several different regions systems. For mammalian studies, the gene might be too long to be conveniently retained in one clone. Therefore a suitable artificial "minigene" will be designed and constructed from suitable parts of the cDNA and genomic clones. The 55 minigene will be under the control of its own promoter or instead will be replaced by an artificial one, perhaps the mouse metallothioneine I

promoter. The resultant 'minigene' will then be 60 introduced into mammalian tissue culture cells e.g. a hepatoma cell line, and selection for clones of cells synthesising maximum amounts of biologically active factor IX will be carried out. Alternatively "genetic farming" could be employed 65 as has been demonstrated for mouse growth hormone (Palmiter et al, Nature 300, 611-615, 1982). The minigene would be micro-injected into the pronucleus of fertilised eggs, followed by in vivo cloning and selection for progeny producing 70 the largest quantity of human factor IX in blood. Alternatively, it is contemplated that the cDNA clone or selected parts of it will be linked to a suitable strong bacterial promotor, e.g. a Lac or

Trp promotor or the lamdba P_R or P_L, and a factor 75 IX polypeptide obtained therefrom. The natural factor IX polypeptide is synthesised as a precursor containing both a signal and propeptide region. They are both normally cleaved off in the production of the definitive length 80 protein. Even this product is merely a precursor. It is biologically inactive and must be gammacarboxylated at 12 specific N-terminal glutamic acid residues in the so called 'GLA' domain by the action of a specific vitamin K-dependent 85 carboxylase. In addition, two carbohydrate molecules are added to the connecting peptide region of the molecule, but is remains unknown whether they are required for activity. The

substrate for the carboxylase is unknown and 90 could be the precursor factor IX polypeptide or alternatively the definitive length protein. Therefore various relevant polypeptides both with and without the precursor domains will be "constructed" using genetic engineering methods

95 in bacterial hosts. They will then be tested as substrates for the conversion of inactive to biologically active factor IV in vitro by the action of partially purified preparations of the carboxylase enzyme which can be isolated from liver 100 microsomes or other suitable sources.

For diagnostic purposes, the recombinant human genomic factor IX DNA or recombinant human mRNA-derived factor IX DNA has a wide variety of uses. It can be cleaved by enzymes or 105 combinations of two or more enzymes into shorter fragments of DNA which can be recombined into the cloning vehicle, producing "sub-clones". These sub-clones can themselves be cleaved by restriction enzymes to DNA molecules suitable for 110 preparing probes. A probe DNA (by definition) is labelled in some way, conveniently radiolabelled, and can be used to examine in detail mutations in the human DNA which ordinarily would produce factor IX. Several different probes have been of the genome where mutation was suspected to have occurred in patients. Failure to obtain hybridisation from such a probe indicates that the

120 In particular it has been shown that Christmas disease can be detected or confirmed by such methodology. Useful probes can contain intron and/or exon regions of the genomic DNA or can

sequence of the probe differs in the patient's DNA.

30

40

45

contain cDNA derived from the mRNA.

The invention includes particularly probe DNA, i.e. which is labelled, and of a length suitable for the probing use envisaged. It can be single-stranded or double-stranded over at least the human factor IX DNA probing sequences thereof and such sequences will usually have a length of at least 15 nucleotides, preferably at least 19—30 nucleotides in order to have a reasonable probability of being unique They will not usually be larger than 5 kb and rarely longer than 10 kb.

The invention accordingly includes a DNA molecule, comprising part of the human factor IX DNA sequence, whether or not labelled, whether intron or exon or partly both. It also includes human cDNA corresponding to part of all of human factor IX mRNA. It includes particularly a solution of any DNA of the invention, which is a form in which it is conveniently obtainable by electroelution from a gel.

The invention includes, of course, a host transformed with any of the recombinant DNA of the invention. The host can be a bacterium, for example an appropriate strain of *E.coli*, chosen according to the nature of the cloning vehicle employed. Useful hosts may include strains of *Pseudomonas*, *Bacillus subtilis* and *Bacillus stearothermophilus*, other *Bacilli*, yeasts and other fungi and mammalian (including human) cells.

One process practised in connection with this invention for preparing a host transformed with the recombinant DNA of the invention is based on the following steps:—

(1) synthesising an oligodeoxynucleotide
35 having a nucleotide sequence comprising that occurring in bovine factor IX messenger RNA coding for amino acids 70—75 or 348—352 of bovine factor IX, and labelling the oligodeoxynucleotide to form a probe;

(2) preparing complementary DNA to a mixture of bovine mRNAs:

(3) Inserting the complementary DNA in a cloning vector to form a mixture of recombinant bovine cDNAs;

(4) transforming a host with said mixture of recombinant bovine cDNAs to form a library of clones and multiplying said clones;

(5) probing the clones with the synthetic oligodeoxynucleotide probe obtained in step 1 and isolating the resultant recombinant bovine factor IX cDNA-containing clone;

(6) digesting the recombinant bovine factor IX cDNA from said clone with one or more enzymes to produce a bovine factor IX cDNA molecule comprising a shorter sequence of bovine factor IX DNA, but preferably at least 50 base-pairs long; and

(7) probing a library of recombinant human genomic DNA in a transformed host with the shorter sequence bovine factor IX cDNA molecule, to hybridise the human genomic DNA to the said recombinant bovine factor IX DNA and isolating the resultant recombinant DNA-transformed host.

Brief description of the drawings

65 Figure 1 shows the structure of a published amino-acid sequence of bovine factor IX polypeptide, the deduced sequence of the mRNA from which it would be translated and the structures of oligonucleotides (oligo-N1 and N2) synthesised in the course of this invention;

Figures 2 and 3 show the chemical formulae of "building blocks" used to synthesise the oligonucleotides referred to in Figures 1 and 11;

Figure 4 is an elevational view, partly sectioned, 75 showing an apparatus for synthesising oligonucleotides;

Figure 5 shows the sequence of part of the bovine factor IX cDNA obtained in this invention;

Figure 6 is a map showing the organisation of 80 an approximately 27 kb length of human factor IX genomic DNA and is divided into five portions, showing:—

(a) the exon regions;

(b) the 11,873-nucleotide length sequenced;

85 (c) cDNA molecules obtained by restriction with various endonucleases, sub-cloned and subsequently used as probes;

(d) DNA molecules obtained by restriction with various endonucleases; and

90 (e) three regions of human factor IX genomic DNA derived from three clones in lambda phage vector.

Figure 7 shows the sequence of the DNA of Figure 6(b) and in parts the encoded protein;

Figure 8 shows a restriction enzyme chart of the sequence shown in Figure 7;

Figure 9 shows part of the sequence of the human factor IX cDNA and its encoded protein;

Figure 10 shows the structure of a pair of complementary oligonucleotides (oligo N3 and N4) synthesised in the course of this invention;

Figure 11 shows part of the DNA sequence of the vector pAT153/Pvull/8 of this invention, in the region where it differs from pAT153;

105 Figure 12 is a diagram of plasmid pHIX17 of the invention showing the origin of the 1.4 kb fragment used for probing and initial sequencing; and

Figure 13 shows the position of the major
110 radioactive bands on probing a "Southern blot" of
normal human DNA, cut by the restriction
enzymes EcoRI(E), HindIII(H), Bg/II(B) and BcII(Bc),
with a sub-clone of the recombinant human factor
IX DNA of this invention.

115 DESCRIPTION OF PREFERRED EMBODIMENTS 1. General description

A recombinant DNA of the invention can be extracted by means of probes from a library of cloned human genomic DNA. This is a known recombinant library and the invention does not, of course, extend to human genomic factor IX DNA when present in such a library. The probes used were of bovine factor IX cDNA (DNA complementary to bovine mRNA), which were

20

25

prepared by an elaborate process involving firstly the preparation of recombinant bovine cDNA from a bovine mRNA starting material, secondly the chemical syntheses of oligonucleotides, thirdly their use to probe the recombinant bovine cDNA, in order to extract bovine factor IX cDNA and fourthly the preparation of suitable probes of shorter length from the recombinant bovine factor IX cDNA. The first probe tried appeared to contain an irrelevant sequence and the second probe tried not containing it, proved successful in enabling a single clone of the human genomic factor IX DNA to be isolated. This clone is designated lambda HIX-1. The steps involved are described in more detail in the sub-section "Examples" appearing hereinafter, and the second probe comprises the 247 base-pair DNA sequence of bovine factor IX cDNA indicated in Figure 5 of the drawings. The invention therefore provides specifically a recombinant DNA which comprises a cloning vehicle sequence and a DNA sequence foreign to the cloning vehicle, which recombinant DNA hybridises to a 247 base-pair sequence of bovine factor IX cDNA indicated in Figure 5 (by the arrows at each end thereof).

The cloning vehicle or vector employed in the invention can be any of those known in the genetic engineering art (but will be chosen to be compatible with the host). They include *E.coli*. plasmids, e.g. pBR322, pAT153 and modifications thereof, plasmids with wider host ranges, e.g. RP4 plasmids specific to other bacterial hosts, phages, especially lambda phage, and cosmids. A cosmid cloning vehicle containing a fragment of phage DNA including its cos (cohesive-end site) inserted in a plasmid. The resultant recombinant DNA is circular and has the capacity to accommodate very large fragments of additional foreign DNA.

Fragments of human factor IX genomic DNA
can be prepared by digesting the cloned DNA with
various restriction enzymes. If desired, the
fragments can be religated to a cloning vehicle to
prepare further recombinant DNA and thereby
obtain "sub-clones". In connection with this
embodiment a new cloning vehicle has been
prepared. This is a modified pAT153 plasmid
prepared by ligating a BamHI and HindIII double
digest of pAT153 to a pair of complementary
double sticky-ended oligonucleotides having a
DNA sequence providing a BamHI restriction
residue at one end, a HindIII restriction residue at
the other end and a PvulI restriction site in
between.

While the invention is described herein with
reference to human genomic factor IX DNA in
particular, the invention includes human factor IX
cDNA (complementary to human factor IX mRNA)
which contains substantially the same sequences.
A library of human cDNA has been prepared and
probed with human factor IX genomic DNA to
isolate human factor IX cDNA from the library. For
this purpose the probe DNA is conveniently of
relatively short length and must include at least
one exon sequence. The invention therefore
includes a process of preparing a host transformed

with recombinant DNA, comprising cloning vector sequences and a sequence of nucleotides comprised in cDNA complementary to human factor IX mRNA, which process comprises probing a library of clones containing recombinant DNA complementary to human mRNA with a probe comprising a labelled DNA comprising a sequence complementary to part or all of an exon region of the human factor IX genome.

75 2. Examples

A. Bacteria used

E.coli K—12 strain MC 1061 (Casadaban & Cohen, J.Mol.Biol. 138, 179—207, 1980), E.coli K—12 strain HB 101 (Boyer & Roulland-Dussoix, J.Mol.Biol 41, 459—472, 1969) and E.coli K—12 strain K803 which is a known strain used by genetic engineers.

B. Source and purification of bovine factor IX, anti bovine factor IX antibody, and bovine mRNA

Highly purified bovine factor IX and rabbit antibovine factor IX antiserum were gifts from Dr. M.
P. Esnouf. Analysis of the purified bovine factor IX
on a denaturating polyacrylamide gel showed that
it has a purity of greater than 99%. Specific antifactor IX immunoglobulins used for
immunoprecipitation experiments were purified as
described by Choo et al., Biochem.J. 199,
527—535, 1981, by passage of the crude
antiserum through a Sepharose—4B column onto
which pure bovine factor IX has been coupled.
Bovine mRNA was obtained from calf liver and

isolated by the guanidine hydrochloride method (Chirgwin et al., Biochem. 18, 5294—5299, 100 1979). The mRNA preparation was passaged through an oligo dT-cellulose column (Caton and Robertson, Nucl. Acids Res. 7, 1445—1456, 1979) to isolate poly(A) + mRNA. Poly(A) + mRNA was translated in a rabbit reticulocyte cell-free system in the presence of 35S-cysteine as described by Pelham and Jackson (Eur. J.Biochem. 67, 247—256, 1976). At the end of the translation reaction, factor IX polypeptide was precipitated by the addition of specific anti-factor IX immunoglobulins. The immunoprecipitation procedure was as described

The immunoprecipitated material was washed throughly and resolved on a two-dimensional SDS-polyacrylamide gel (Choo et al., Biochem.J. 181, 285—294, 1979), by isoelectric focussing in one dimension and electrophoresis in another. Some polypeptides of known molecular weight were subjected to this procedure, to serve as

by Choo et al., Biochem.J. 181, 285-294, 1979.

120 reference points. The immunoprecipitated material showed 4 pronounced spots, all in the 50,000 molecular weight region and with separated isoelectric points. These predominant spots of molecular weight about 50,000 represent a single polypeptide chain plus a possible prepeptide

125 polypeptide chain plus a possible prepeptide signal sequence, a deduction compatible with published data (Katayama et al., Proc. Natl.Acad. Sci.USA 76, 4990—4994, 1979).

When the gel analysis was repeated for the

same material but immunoprecipitated in the presence of unlabelled pure bovine factor IX, the 4 spots appeared at reduced intensity, indicating that the translation product is specifically competed for by pure factor IX. Thirdly, immunoprecipitation was performed using a control rabbit antiserum, i.e. from a rabbit which had not been immunised with factor IX. None of the 4 spots appeared. These results therefore 10 indicate that the translation product was a factor IX polypeptide.

The specific immunological/cell-free translation assay established above was used to monitor the enrichment of factor IX mRNA on sucrose gradient 15 centrifugations. Total poly(A) + mRNA was resolved by two successive separations by sucrose gradient centrifugations. When individual fractions from the gradient were assayed by the above method, a fraction of size 20-22 Svedberg units 20 (approx. 2.5 kilobases of RNA) region was found to be enriched (approx. ten-fold) for the bovine factor IX mRNA. This enriched fraction was used in the subsequent cloning experiments.

25 C. Synthesis of specific bovine factor IX deoxyoligonucleotide mixtures

Starting from a knowledge of the amino acid sequence of bovine factor IX (Katayama et al., Proc.Natl.Acad.Sci. USA 76, 4990-4994, 30 1979), the synthesis of two mixtures of oligonucleotide probes was designed. These probes consisted of DNA sequences coding for two different regions of the protein. The regions selected were those known to differ in sequence in 35 the analogous serine proteases, prothrombin, Factor C and Factors VII and X and were those corresponding to amino acids 70-75 and 348—352 respectively. The 70—75 region was particularly favourable in that the mixture of 40 oligonucleotides synthesised, i.e. oligo N2A and oligo N2B, contained all 16 possible sequences that might occur in a 17 nucleotide long region of

the mRNA corresponding to amino acids 70-75.

The oligo N2A-N2B mixture is hereinafter called

Figure 1 of the drawings shows the two selected regions of the known amino acid sequence of bovine factor IX, the corresponding mRNA and the oligonucleotides synthesised. 50 Since some of the amino acids are coded for by more than one nucleotide triplet, there are 4 ambiguities in the mRNA sequence shown for amino acids 70-75 and therefore 16 possible

individual sequences.

45 "oligo N2" for brevity.

The nucleotide mixtures oligo N1 and oligo N2 55 were synthesized using the solid phase phosphotriester method of Duckworth et al.. Nucl.Acids Res. 9, 1691-1706, 1981, modified in two ways. Firstly, o-chlorophenyl rather than p-60 chlorophenyl blocking groups were used for the phosphotriester grouping, and were incorporated in the mononucleotide and dinucleotide "building blocks". Figures 2 and 3 of the drawings show (a) dinucleotide and (b) mononucleotide "building 65 blocks". DMT = 4.4' - dimethoxytrityl and B = 6-

N-benzoyl-adenin-9-yl, 4-N-benzoylcytosin-1-yl, 2-N-isobutyrylguanin-9-yl or thymin-1-yl, depending on the nucleotide selected. Secondly, the "reaction cell" used for the successive 70 addition of mono- or dinucleotide "building blocks" was miniaturised so that the coupling step with the condensing agent 1-(mesitylene-2sulphonyl)-3-nitro-1,2,4-triazole (MSNT) was carried out in a volume of 0.5ml pyridine containing 3.5 micromoles of polydimethylacrylamide resin, 17.5 micromoles of incoming dinucleotide (or 35 micromoles of mononucleotide) and 210 micromoles of MSNT.

Figure 4 of the drawings is an elevational view of the microreaction cell 1 and stopper 2 used for 80 oligonucleotide synthesis, drawn 70% of actual size. The device comprises a glass-to-PTFE tubing joint 3 at the inlet end of the stopper 2. The stopper has an internal conduit 4 which at its 85 lower end passes into a hollow tapered ground glass male member 5 and thence into a sintered glass outlet 6 to the stopper. The cell 1 has a

ground glass female member 7 complementary to the member 5 of the stopper, leading to reaction chamber 8, the lower end of which terminates in a sintered glass outlet 9. This communicates with glass tubing 10 and a 1.2mm. "Interflow" tap 11. Further glass tubing 10, beyond the tap 11, leads to the outlet glass-to-PTFE tubing joint 12. Pairs of ears 13 on the stopper and cell enable them to be joined together by springs (not shown) in a 95 liquid-tight manner.

After completion of the synthesis and deprotection, fractionation was carried out by high pressure liquid chromatography (Duckworth et al., see above) and the peak tubes corresponding to 100 the product of correct chain length were located by labelling of fractions at their 5'-hydroxyl ends using [gamma-32p]-ATP and T4 polynucleotide kinase, followed by 20% 7M urea polyacrylamide gel electrophoresis. The position on the gel of the 17- and 14- oligonucleotides was determined by separately labelling, by the method described above, 17- and 14- nucleotide long "marker" oligonucleotides and subjecting these to the same gel electrophoresis.

110 D. Preparation of libraries of cDNA sequences for bovine mRNA

> Two different approaches were used for the generation of cloned cDNA library:-

(i) Mbol library First strand cDNA was 115 synthesised using the sucrose gradient-enriched poly(A)+bovine mRNA as template. The conditions used were as described by Huddleston & Brownlee, Nucl. Acids Res. 10, 1029-1030,

120 1981, except that 2 micrograms of oligo N-1, 20-30 micrograms of the mRNA, 10 microcuries [alpha-32P]-dATP (Amersham, 3000 Ci/mmole), and 50 U of reverse transcriptase were used in a 50 microlitre reaction. "dNTP" in Figure 1 denotes

125 the mixture of 4 deoxynucleoside triphosphates required for synthesis. Oligo N-1 hybridises to the corresponding region on the mRNA (refer to Figure 1) and thereby acts as a primer for the

initiation of transcription. It was used in order to achieve a further enrichment for factor IX mRNA. At the end of the cDNA synthesis reaction, the cDNA was extracted with phenol and desalted on a Sephadex-G100 column, before it was treated with alkali (0.1 M NaOH, 1 mM EDTA) for 30 min. at 60°C to remove the mRNA strand. Second strand DNA synthesis was then carried out exactly as published (Huddleston & Brownlee, Nucl.Acids Res. 10, 1029—1038, 1981).

The double-stranded DNA was next cleaved with the restriction enzyme *Mbol* and ligated to the plasmid vector pBR322 which had been cut with *Bam*HI and treated with calf intestinal alkaline phosphatase to minimise vector self-religation. Phosphatase treatment was carried out by incubating 5 micrograms of *Bam*HI-cut pBR 322 plasmid with 0.5 microgram calf intestinal phosphatase (Boehringer; in 10mM Tris — HCI buffer, pH 8.0) in a volume of 50 microlitres at 37°C for 10 minutes, see Huddleston & Brownlee

The ligated DNA was used to transform *E.coli* strain MC 1061. For transformation *E.coli* MC 1061 was grown to early exponential phase as indicated by an absorbancy of 0.2 at 600 nm and made "competent" by treating the pelleted bacterial cells first with one half volume, followed by repelleting, and then with 1/50 volume of the original growth medium of 100mM CaCl₂ 15% v/v glycerol and 10mM PIPES—NaOH, pH 6.6 at 0°C. Cells were immediately frozen in a dry ice/ethanol bath to -70°C. For transformation, 200 microlitre

aliquots were mixed with 10 microlitres of the recombinant DNA and incubated at 0°C for 10 minutes followed by 37°C for 5 minutes. 200 microlitres of L-broth (bactotryptone 10g., yeast extract 5g., sodium chloride 10g., made up to 1 litre with deionised water) were then added and

322. Of these, aprox. 85% were found to be tetracycline-sensitive.

(ii) dC/dG tailed library In the preparation of this library, first strand cDNA was synthesised as 50 described for the above library except that oligo dT(12-18) was used as a primer to initiate cDNA synthesis. Following this, the cDNA was tailed with dCTP using terminal transferase and backcopied with the aid of oligo dG₍₁₂₋₁₈₎ primer and 55 reverse transcriptase to give double stranded DNA, exactly according to the method of Land et al., Nucl.Acids Res. 9, 2251-2266, 1981. After a further tailing with dCTP, this material was annealed by hybridisation to a dGTP-tailed 60 pBR322 plasmid at the Pstl site. The hybrid DNA was used to transform E.coli strain MC 1061. A library of approximately 10,000 tetracyclineresistant colonies was obtained. Of these, approximately 80% were found to be sensitive to

65 ampicillin, due to insertion of DNA into the

ampicillin-resistant gene at the Pstl site.

E. Isolation of specific bovine factor IX clones (i) From Mbol library

The library of colonies, in an unordered fashion, was transferred onto 13 Whatman 541 filter papers and amplified with chloramphenicol, to increase the number of copies of the plasmid in the colonies, as described by Gergen et al., Nucl. Acids Res., 1, 2115—2136 (1979). The filters

75 were pre-hybridised at 65°C for 4h in 6 x NET (1xNET = 0.15m NaCl, 1mM EDTA, 15mM Tris-HCl, pH 7.5), 5 x Denhardt's, 0.5% NP40 non-ionic surfactant, and 1 microgram/ml. yeast RNA as described by Wallace et al., Nucl. Acids Res. 9,

80 879—894 (1981). Hybridisation was carried out at 47°C for 20h in the same solution containing 3 x 10⁵cpm (0.7 nanogram/ml) of labelled oligo N—2 probe. Labelling was done by phosphorylation of the oligonucleotides at the 5'

hydroxyl end using [gamma-32P]-ATP and T4 phophokinase (Huddleston & Brownlee, Nucl.Acids Res. 10, 1029—1038, 1981). At the end of the hybridisation, filters were washed successively at 0—4°C (2h), 25°C (10 min),

90 37°C (10 min) and 47°C (10 min). After radioautography of the filters from this screening, one colony showed a positive signal above background. This colony was designated BIX—1 clone.

95 (ii) From dC/dG-tailed library

Screening of this library, in an ordered array fashion, using oligo N—2 probe as described above has resulted in the identification of a positive clone. This was designated BIX—2 clone

100 F. Sequence characterisation of bovine factor IX cDNA clones

CDNA clones
Characterisation of BIX—1 clone by restriction

endonuclease cleavage indicated that it contained a DNA insert of about 430 base-pairs (data 105 omitted, for brevity). Figure 5 shows part of the nucleotide sequence of the coding strand, determined by the Maxam-Gilbert method, extending over 304 nucleotides and provides direct evidence that it has the identity of a bovine

110 factor IX sequence. Thus, nearly all of this 304 nucleotide sequence (corresponding to the amino acid residues 52—139) agrees with the nucleotide sequence predicted from the known bovine factor IX amino acid sequence data

115 (Katayama et al., Proc.Natl.Acad.Sci. 76, 4990—4994, 1979). Over this region, there are no discrepancies between BIX—1 and these published data for factor IX, except at nucleotides 38—40 where the amino acid coded for is Asp

120 instead of Thr. This amino acid change was similarly observed in a second, independent cDNA clone (BIX—2; see below). The remainder of the 304-nucleotide sequence, i.e. that shown in brackets in Figure 5, does not agree with the
 125 published bovine factor IX amino acid data of

Katayama.

In Figure 5, the underlined portion denotes the sequence corresponding to the oligo N—2 probe

sequence, the asterisk denotes a nonsense codon, the brackets enclose a sequence which does not correspond to Katayama's amino acid data and the arrows indicate *Hinfl* restriction sites. The Katayama numbering system for amino acids is shown and this sequence is in the opposite orientation to the direction of transcription of the tetracycline-resistant gene of the plasmid.

By similar methods, BIX—2 clone was found to have a DNA insert of 102 nucleotides and this spans the nucleotide positions 7—108 as shown in Figure 5. The nucleotide sequences for BIX—1 and BIX—2 clones over this region (nucleotide 7—108) were identical.

15

G. Isolation of human factor IX gene (i) *Initial clone — Iambda HIX—1*

A library of cloned human genomic DNA, namely a Haelll/Alul lambda phage Charon 4A
library prepared by Lawn et al., Cell, 15,
1157—1174, 1978, was used. 106 phage recombinants from this library were screened using the in situ plaque hybridisation procedure as described by T. Maniatis et al., Cell, 15, 687,
1978. Pre-hybridisation and hybridisation were carried out at 42°C in 50% formamide. After hybridisation, filters were washed at room temperature with 2 x SSC (1 x SSC = 0.15mM NaCl, 15mM sodium citrate, at pH 7.2) and 0.1% SDS, then at 65°C with 1 x SSC and 0.1% SDS.

Two DNA molecules, being restriction fragments from the factor IX cDNA cloned in BIX—1, were radiolabelled and used as probes in the hybridisation. The first fragment corresponds to nucleotide numbers —8 to 317 on the numbering system of Figure 5, and was isolated by Sau3Al digestion of BIX—1 plasmid DNA. The isolated DNA was labelled to high specific activity by incorporation of [alpha—3²P] -dATP using a nick translation (Rigby et al., J. Mol.Biol. 113, 237—251, 1977, modified, vide infra). Using this probe, 10 clones were isolated. These were plaque-purified and re-hybridised with a 247-nucleotide fragment from BIX—1 clone. This fragment, derived from nucleotides 3—249 can be seen from Figure 5. It contains only sequences

be seen from Figure 5. It contains only sequences in agreement with the Katayama bovine factor IX amino acid sequence and was isolated by *Hinfl* digestion of BIX—1 plasmid DNA. Only a single clone gave a positive hybridisation signal with this 247-nucleotide probe. This clone was further plaque-purified and the resulting clone was designated "lambda HIX—1".

(ii) Subsequent genomic clones

A sub-clone, pATIXcVII, of recombinant human factor IX cDNA from human liver mRNA, and prepared as described in Section L below, was linearised by digestion with HindIII and BamHI.

The resulting 2 kb cDNA molecule was purified by 1% agarose gel electrophoresis. After electroelution, about 100 ng of this cDNA was nick-translated with [alpha 32p] dATP (see above) and used as a hybridisation probe to screen the HaelII/AIuI lambda phage Charon 4A human

genomic DNA library for further genomic clones.

using standard stringent hybridisation conditions. Two further human factor IX genomic clones, designated lambda HIX—2 and lambda HIX—3, were thus obtained.

70 H. Characterisation of human factor IX genomic clones

(i) Restriction map

The initial lambda HIX—1 clone was characterised by cleavage with various single and double digests with different restriction endonucleases and Southern blotting of fragments using the bovine factor IX cDNA probe (results omitted for brevity). The subsequently isolated lambda HIX—2 and 3 clones were characterised in the same way except that the human cDNA

probe, pATIXcVII (see Section L below) was used for the Southern blots. From these results it emerged that the sequences in the factor IX genome corresponding to lambda HIX—2 and 3

85 overlapped with lambda HIX—1 as shown in Figure 6(e). In Section (d) of this Figure 6 are summarised the results of the analysis using the restriction enzymes EcoRI (E), HindIII (H), Bg/II (B), BamHI (Ba) and PvuII (P), and this serves as a
 90 restriction enzyme map.

restriction enzyme ma (ii) Sequencing

'w" exon.

Numerous sub-clones were isolated from a knowledge of the rectriction enzyme map as described in Section J(ii) below, the majority in a vector pAT153/Pvull/8. Examples of these sub-clones are shown in Figure 6(c) and a number were used and were of a convenient length for sequence analysis by the Maxam-Gilbert method (Maxam & Gilbert, Proc.Natl.Acad.Sci.USA 74,

56—564, 1980).

Initially sequencing was done on part of a 1.4 kb *EcoR*I restriction fragment from the sub-clone pHIX—17, see below and J(i). A 403-nucleotide (base-pair) length was sequenced, of which a 129-nucleotide length was identified as lying

15 129-nucleotide length was identified as lying within an exon region. This is the 129-nucleotide sequence used above to define the factor IX DNA. Subsequently, a region of 11873 bases was

sequenced in the central portion of the gene [see Figure 6(b)]. Figure 7 shows the sequence of one strand of the DNA. The nucleotides are arbitrarily numbered from 1 to 11873 in the 5' to 3' direction. The original 403-nucleotide sequence runs from Figure 7 nucleotides Nos. 4372 to 4774 and is indicated by O—O'. The 129-nucleotide sequence lying within the 403 one, runs from Figure 7 nucleotides Nos. 4442 to 4570 and is indicated by J—J'. This corresponds exactly to the

120 In detail, the sequence of nucleotides Nos.
1—7830 contains two short exons (nucleotides
4442—4570 and 7140—7342 respectively)
marked w and x in Figure 6(a), J—J' and J'—J'' in
Figures 7 and 9. These code for amino acids

125 85—127, and 128—195 respectively of the amino acid sequence predicted from the human factor IX cDNA clone (Figure 9). There are no differences in amino acid sequences predicted from the genomic and cDNA clones of the

invention in these two exon regions. The sequence of the gene between residues 7831—11873 is less complete, containing several gaps, but is still a useful characterisation of the gene as it contains two "Alul repeat" sequences, nucleotides 7960—8155 and 9671—9938. Alul sequences are found in many genes. The repetition is not exact but there is a typical degree of homology between them. This further characterisation provides a useful cross-check on the accuracy of the restriction enzyme map. This emerges more clearly from the restriction enzyme chart of Figure 8.

Figure 8 is a chart produced by a computer 15 analysis of the sequence data of the 11873 nucleotide long sequence of Figure 7. Column 1 of Figure 8 gives the arbitrary nucleotide number allotted to the nucleotide of Figure 7. Column 2 apportions the nucleotide number as a fraction of 20 the whole sequence. Column 3 shows the restriction enzymes which will cut the DNA within various short sequences of nucleotides shown in Column 4. The short sequences of Column 4 begin with the nucleotide numbered in Column 1. With 25 the aid of this chart the positions of the restriction sites shown in Figure 6(d) and some of the sequences shown in Figure 6(c) can be determined very accurately. For example sequences II-IV are produced by restriction at the following sites (denoted by the first nucleotide 30 number at the 5' end of each site).

> II 3624 — 4769 III 6380 — 7378 IV 10589 — 11868

Particularly important sites are arrowed in Figure 8. Some of the relevant nucleotide numbers are shown in Figure 6(c), the number given being that of the nucleotide at the 5' end of each site.

Further sequence analysis of the sub-clones V,
VI, VII and VIII shown in Figure 6(c) indicates that
the factor IX gene is divided into at least 7 exon
regions separated by at least 6 introns. The
positions of the exons are shown in Figure 6(a) by
the solid blocks labelled t, u, v, w, x, y and z. The
"z" exon is much the longest and its 3'-end
coincides with the 3'-end of the mRNA. The
location of these exons relative to the cDNA
sequence is discussed below (section L) and it is
clear that the "t" exon shown in Figure 6(a) is not
a marker for the 5'-end of the gene, as its
sequence fails to match that of the extreme 5'-end
of the cDNA clone (see below). This suggests that
the factor IX gene will be longer at its 5'-end than

Additionally, pHIX—17 DNA was digested with *Eco*RI. The digested material was resolved on 0.8% agarose gel and a 1.4 kb fragment was isolated in solution by electroelution. It can be stored in the usual manner. This 1.4 kb long molecule was used for the initial sequencing. Only about 1.0 kb is inserted DNA, the remaining 0.4 kb being of pBR322. A 403 nucleotide length of the

the 27 kb region shown in Figure 6, and will

contain at least one further exon.

inserted DNA was sequenced and is identified as 0—0' in Figure 7. The same 1.4 kb fragment was also labelled and used as a probe in Section M.

I. Construction of a vector pAT153/Pvull/8 A derivative of the plasmid pAT153 (Twig & Sherratt, Nature 283, 216—218, 1980) was prepared for subcloning of Pvull fragments of factor IX genomic clones, and for ease of characterisation of the resultant subclones. Two partially complementary synthetic deoxyoligonucleotides, oligo N3, and, oligo N4, 75 were synthesised by the solid phase phosphotriester method described in Section C above. Each has "overhanging" BamHI and HindIII recognition sequences and an internal Pvull recognition sequence. Figure 10 shows the 80 structures of oligo N3 and oligo N4. BamHI and HindIII cleave ds DNA to leave sticky or "overhanging" ends. For example HindIII cleaves

- --- AAGCTT
- TTCGAA

between the adenine-carrying nucleotides of each strand leaving the sticky-ended complementary strands:—

— A — TTCGA

90 which are present in the oligo N3/N4 combination. pAT153 was digested with HindIII and BamHI and the 3393 nucleotide long linear fragment was separated from the 346 nucleotide shorter fragment by 0.7% agarose gel electrophoresis, 95 followed by electroelution of the appropriate bands visualised by ethidium bromide fluorescence under UV light. After treatment with calf intestinal phosphatase, as described in Section D(i), the BamHI-Hind III 3393-long 100 fragment was ligated to an equimolar mixture of oligo N3 and oligo N4 which themselves had been pretreated, as a mixture, with T4 polynucleotide kinase and ATP, to phosphorylate their respective 5'-terminal OH groups. After transforming 105 competent MC 1061 cells (see above) and plating on L-broth plates containing 20 micrograms/ml final concentration of ampicillin, 11 colonies were selected for further analysis. 1 ml plasmid preparation, see Holmes and Quigley, Analytical 110 Biochem. 114, 193-197 (1981), was isolated from the 11 colonies. The plasmid DNA was then analysed for its ability to be linearised by the restriction enzymes BamHI, HindIII and Pvull. Four clones were positive in this assay and one, 115 labelled pAT153/Pvull/8, was selected for sequence analysis by the Maxam-Gilbert method across the newly constructed section of the plasmid. This part of the sequence is shown: Figure 11 along the unique restriction sites. The 120 novel part of the plasmid sequence is underlined: the remainder is present in the parent plasmid pAT153. The vector allows blunt-end cloning

(after treatment with phosphatase) into the

inserted Pvull site. The cloned DNA can be excised, assuming that it lacks appropriate internal restriction sites, with BamHI/HindIII, BamHI/ClaI or BamHI/EcoRI double digests. The sites adjacent to the Pvull site are also convenient for end labelling with 32P for characterization of the ends of cloned DNA by the Maxam-Gilbert sequencing method.

J. Sub-cloning of human factor IX gene

The following subcloning experiments were 10 carried out as a first step towards sequencing of the factor IX gene, and to facilitate the isolation of a small DNA fragment to be used as a probe for the analysis of genomic DNA from haemophilia B patients (see sections M).

15 (i) Sub-cloning into pBR322 plasmid

An approximately 11 kilobase Bg/II fragment (see Figure 6) within the factor IX DNA insert in lambda HIX-1 clone was inserted into the BamHI site of pBR322. Transformation was carried out in 20 the E.coli strain, HB 101. The resulting "subclone" was designated pHIX-17 (Figure 12). (i) Sub-cloning into pAT153/PvuII/8 (a) Plasmid DNA from pHIX—17 was prepared and cleaved with Pvull. Five discrete fragments, all

25 derived from the DNA insert of pHIX-17, were isolated. The sizes of these fragments were approximately 2.3, 1.3, 1.2, 1.1 and 1.0 kilobases. These fragments were blunt-end ligated into the Pvull site of the pAT153/Pvull/8 vector and

30 transformed into E.coli HB 101. Five clones of recombinant DNA which carried the 2.3, 1.3, 1.2, 1.1 and 1.0 kb fragments were obtained and these were designated pATIXPvu-1, 2, 3, 4 and 5 respectively. Factor IX DNA from pATIXPvu-2 is 35 abbreviated as IV and pATIXPvu-5 as III in Figure 6(c).

(b) Phage DNA from the lambda HIX—1 genomic clone was digested with EcoRI. Three different fragments (approximately 5, 2.3, 0.96, kb; see 40 Figure 6), all derived from the insert into the phage, were isolated and inserted in pAT153/Pvull/8 vector at the EcoRI site and cloned in E.coli HB 101 to form sub-clones. The three resulting clones for each of these fragments

45 were designated pATIXEco-1, 2 and 4 respectively which are shown in the restriction map of Figure 6(d). pATIXEco-1 was further digested with both EcoRI and Bg/II, and the "overhanging ends" of the restriction sites filled in with deoxynucleotide

50 triphosphates using the Klenow fragment of DNA polymerase I. After isolation of the resulting 1.1 kb fragment by agarose gel electrophoresis and electroelution, it was blunt-end ligated using T4 DNA ligase into the Pvull site of pAT153/Pvull and

55 allowed to transform E.coli MC 1061. The resultant sub-clone was designated pATIXBE and the factor IX DNA sequence thereof is abbreviated as II in Figure 6(c).

(c) Phage DNA from lambda HIX-2 was 60 digested with HindIII and EcoRI giving a 1.8 kb and a 2.6 kb fragment amongst others. These fragments were eluted separately, filled in as described in (b) above, cloned as above into the Pvull site of pAT153/Pvull/8 and allowed to

65 transform E.coli MC 1061. The resultant clones were designated pATIXHE-1, and the factor IX DNA sequence thereof is abbreviated as V in Figure 6(c), and pATIXEco—6 and the factor IX DNA sequence thereof is abbreviated as VI in Figure 6(c).

(d) Phage DNA from lambda HIX-3 was digested with EcoRI and Hind III and the fragments of 2.3 kb and 2.7 kb were sub-cloned exactly as described in (c) above. The resultant clones were designated pATIXEH-1, abbreviation VII in Figure 6(c), and pATIXHE—2, abbreviation VIII in Figure 6(c).

K. Preparation of a library of cDNA clones from 80 human liver mRNA

Messenger RNA was extracted from a human liver and a 20—22 Svedberg unit enriched fraction of mRNA prepared exactly as described for bovine mRNA in Section B above, except that a

85 'translation assay' was not used. The first steps in the construction of the double-stranded DNA were carried out using the 'Stanford protocol' kindly supplied from Professor P Berg's department at Stanford University, USA. This itself is a modification of Wickens, Buell & Schimke

(J.Biol.Chem. 253, 2483—2495, 1978) and some further modifications, incorporated in the description given below were made in the present work.

95 For the first strand cDNA synthesis 6 micrograms of poly(A)+ 20-22S human mRNA was incubated with 5 microlitres of 10x buffer (0.5 M Tris-chloride, pH 8.5 at room temperature, 0.4 M KCI, 0.008M MgCl, and 4 mM

100 dithiothreitol), 20 microlitres of a 2.5 mM mixture of each of the four deoxynucleoside triphosphates, 0.5 microlitres of oligo $dT_{(12-18)}$, 1 microlitre (containing 0.5 microcurie) of [alpha- 32 P] -dATP, 2 microlitres of reverse transcriptase (14 units per

105 microlitre) and the volume made up to 50 microlitres with deionized water. After incubation for 1 hour at 42°C, the solution was boiled for 1\frac{1}{3} minutes and then rapidly cooled on ice. The second strand synthesis was carried out by adding

110 directly to the above solution 20 microlitres of 5x second strand buffer (250 mM Hepes/KOH pH 6.9, 250 mM KCl, 50mM MgCl₂), 4 microlitres of a 2.5 mM mixture of each of the four deoxynucleoside triphosphates, 10 microlitres of

115 E.coli DNA polymerase I (6 units per microlitre) and making the volume of the solution up to 100 microlitres with deionized water. After incubation for 5 hours at 15°C, S, nuclease digestion was carried out by addition of 400 microlitres of S.

120 nuclease buffer (0.03 M sodium acetate pH 4.4 0.25 M NaCl, 1 mM ZnSO₄) and 1 microlitre of S₄ nuclease (at 500 units per microlitre). After incubating for 30 minutes at 37°C, 10 microlitres of 0.5M EDTA (pH 8.0) was added. Double

125 stranded DNA was deproteinised by shaking with an equal volume of a phenol: chloroform (1:1) mixture, followed by ether extraction of the aqueous phase and precipitation of ds DNA by addition of 2 volumes of ethanol. After 16 hours at -20°C, ds DNA was recovered by centrifugation. DNA polymerase 1 "fill in" of S₁ ends was carried out by a further incubation of the sample dissolved in 25 microlitres of 50 mM tris-chloride, pH 7.5,
5 10 mM MgCl₂, 5 mM dithiothreitol and containing 0.02 mM dNTP and 6 units of DNA polymerase I. After incubating for 10 minutes at room temperature, 5 microlitres of EDTA (0.1 M at pH 7.4) and 3 microlitres of 5% sodium dodecyl sulphate (SDS) were added.

The following part of the protocol differs from the 'Stanford protocol'. The sample was fractionated on a "mini"-Sephacryl S400 column run in a disposable 1 ml pipette in 0.2 M NaCl, 10 15 mM tris-chloride, pH 7.5 and 1 mM EDTA. The first 70% of the "break-through" peak of radioactivity was pooled (0.4 ml) and deproteinised by shaking with an equal volume of n-butanol:chloroform (1:4). To the aqueous phase 20 was added 1 microgram of yeast RNA (BDH) as carrier followed by 2 volumes of ethanol. After 16 hours at --20°C double stranded DNA was recovered by centrifugation for blunt-end ligation into calf intestinal phosphatase-treated Pvull-cut pAt153/Pvull/8, using T4 DNA ligase (see I and J(ii) above). After performing a trial experiment, it was found that when the bulk of the sample was incubated with 200 nanograms of vector DNA in a suitable buffer (1 mM ATP, 50 mM Tris-chloride, pH 7.4, 10 mM MgCl₂ and 12 mM dithiothreitol) and using 10 microlitres of T4 DNA ligase in a total volume of 0.2 ml, then on subsequent transformation of competent E.coli MC 1061 cells a total of 58,000 ampicillin-resistant colonies 35 were obtained. Up to 20% of these were estimated to derive from "background" nonrecombinants derived by religation of the vector itself. This 20-22S cDNA library was amplified by growing the E.coli for a further 6 hours at 37°C. 40 1 ml aliquots of this amplified library were stored at

L. Isolation and sequence analysis of human factor IX cDNA clones

-20°C in L broth containing 15% glycerol, before

screening for factor IX cDNA clones.

45 6000 colonies of the amplified 20-22S human cDNA library were plated on each of ten 15 cm agar plates and after growing overnight were blotted into Whatman 541 filter paper. After preparing filters for hybridisation as described in 50 section E(i) above, the immobilised colonies were probed with a 1.1 kb molecule of [alpha-32P] -nick translated human factor IX genomic DNA isolated from the pATIXBE subclone (Section J, above). This linear 1.1 kb section of factor IX genomic 55 cDNA was isolated from pATIXBE by cleavage with the restriction enzymes BamHI and HindIII, followed by separation of the 1.1 kb section from the vector by 1.5% agarose gel electrophoresis. After electroelution, nick-translation was carried out as 60 before and the material used in a hybridisation reaction for 16 hours at 65°C in 3x SSC, 10x Denhardts solution, 0.1% SDS and 50 micrograms/ml sonicated denatured E.coli DNA

and 100 micrograms/ml of sonicated denatured

herring sperm DNA. After hybridisation filters were washed at 65°C successively in 3x SSC, 0.1% SDS (2 changes, half an hour each) and 2x SSC, 0.1% SDS (2 changes, half an hour each). After radioautography, 7 clones were selected as positive, but on dilution followed by re-screening by hybridisation as above, only 5 proved to be positive. Plasmid DNA was isolated from each of these 5 clones and one, designated pATIXcVII, was selected for sequence analysis as it appeared 75 to be the longest of the 5 clones as judged by its electrophoretic mobility on 1% agarose gel electrophoresis. A second shorter clone, designated pATIXcVII was also selected for partial sequence analysis. Sequencing was carried out by the Maxam-80

Gilbert method and a 2778 nucleotide long section of sequence is shown in Figure 9. Nucleotides 115-2002 were derived by sequencing clone pATIXcVII. (The actual extent of this clone is greater as it extends in a 5' direction to nucleotide 17. The sequence between 17 and 111 is inverted with respect to the remainder of the sequence presumably due to a cloning artefact.) Nucleotides 1-130 were derived from clone pATIXcVI which extends from nucleotides 90 1—1548 of Figure 9. The sequence from Nos. 2002-2778 was derived by isolating 4 additional clones designated pATIX108.1 pATIX108.2, pATIX108.3 and pATIXDB. The first 3 were derived from a mini-library (designated GGB108) of the cDNA clones constructed exactly as described in section K above except that sucrose density gradient centrifugation was used instead of chromatography on "Sephacryl" 100 S—400 to fractionate the double-stranded DNA

according to size. A fraction of m.w. from 1 kb—5 kb was selected and an amplified library of 10,000 independent clones containing approximately 20% background non-recombinant clones was obtained. Clone pATIXDB derived from another cDNA library (designated DB1) constructed as described in section K except that total poly A+ · human liver mRNA was used as the starting material and sucrose density gradient centrifugation was used to fractionate the DNA

mini-library GGB 108. The complexity of this library was 95,000 with an estimated background of non-recombinants of 50%. Clones pATIX108.1 and pATIX108.2 were selected from a group of 30 hybridization-positive clones isolated by Grunstein-Hogness screening of the mini library GGB 108 using a ³²P-nick translated probe derived from a Sau3Al restriction enzyme fragment, itself

according to size as in the construction of the

derived from nucleotides 1796—2002 of clone pATIXcVII. From pATIX108.1 the sequence of nucleotides 2009—2756 was determined (Figure 9). Following this the sequence of a part of pATIX108.2, specifically nucleotides

125 1950—2086, provided the overlap with pATIXcVII. The remaining 28 hybridization positive clones were screened by carrying out a triple enzymatic digestion with the restriction enzymes EcoRI, BamHI and HindIII and screening

the product of the digest for an EcoRI restriction fragment extending in the 3' direction from the cut at position 2480. By this approach, clone pATIX108.3 was selected and sequenced from nucleotides 2642-2778. This clone was followed by three A nucleotides, which sequence was confirmed as a vestigial marker for the poly A tail, by the subsequent isolation of clone pATIXDB by a similar method, pATIXDB was sequenced 10 from Nos. 2760—2778 and ended in 42 A nucleotides, thus marking the 3' end of the mRNA.

Figure 9 shows that the predicted amino acid sequence codes codes for a protein of 456 amino acids, but included in this are 41 residues of 15 precursor amino acid sequence preceding the N-

terminal tyrosine residue (Y) of the definitive length factor IX protein. The precursor section of the protein shows a basic amino acid domain (amino acids -1 to -4) as well as the more usual 20 hydrophobic signal peptide domain (amino acids -21 to -36).

The definitive factor IX protein consists of 415 amino acids with 12 potential gammacarboxyglutamic acid residues at amino acids 7, 8, 25 15, 17, 20, 21, 26, 27, 30, 33, 36 and 40. Two potential carbohydrate attachment sites occur at amino acid residues 157 and 167. The activation peptide encompasses residues 146—180, which are cut out in the activation of Factor IX (see 30 Background of Invention) by the peptide cleavage

of an R-A and R-V bond. This leaves a light chain spanning residues 1-145 and a heavy chain spanning residues 181-415.

The exact location of the boundaries between 35 exons (see Section H, above) and how they are joined in the mRNA is marked in Figure 9. The exons are marked t, u, v, w, x, y, z. It can be seen that there is a rough agreement between the exon domains and the protein regions. For example, the exon for the signal peptide is distinct from that of the GLA region. Also that of the activation peptide is separated from the serine protease domain.

The 3' non-coding region of the mRNA is extensive, consisting of 1390 residues (including the UAAUGA double terminator 1389—1394 but excluding the poly A tail).

The factor IX cDNA is cleavable by the restriction enzyme HaellI to give a fragment from nucleotides 133-1440 i.e. a 1307 nucleotide 50 long region of DNA entirely encompassing the definitive factor IX protein sequence. The nucleotide sequence recognised by HaellI is GGCC. This fragment should be a suitable starting material for the expression of factor IX protein from suitable promoters in bacterial, yeast of mammalian cells. Another suitable fragment could be derived using the unique Stul site at residue 41 (corresponding to an early part of the hydrophobic signal peptide region) and linking it to a suitable promoter. The nucleotide sequence recognised by Stul is AGGCCT

M. Southern Analysis of normal and patient Christmas disease DNA

(i) Normal

65 The standard (Southern) blotting procedure. Southern, J.Mol. Biol. 98, 503-517, 1975) was used. In a typical experiment, 10-20 micrograms of human genomic DNA (prepared from uncultured blood cells or cultured lymphocytic cells) were digested with one of a number of restriction endonucleases and loaded onto a single gel slot. Following electrophoresis on 0.8% agarose gel and transfer onto nitrocellulose it was hybridised with a probe of 32P-labelled probe II or 75 of 1.4 kb EcoRI fragment (see Section H). Labelling of the probe was carried out by nick translation using the method of Rigby et al., supra, modified as follows. About 100 nanograms of the

probe was mixed with 40 microcuries of lalpha 80 32Pl dATP (activity about 3,000 Curies/mMole, obtained from Amersham International PLC) in 0.05M Tris-HCl, pH 7.5, 0.01M MgCl₂, 0.001M dithiothreitol and dCTP, dGTP, dTTP each at a final concentration of 20 micromolar in a volume of 29 microlitres. To this was added 1 microlitre of "solution X" made up of a mixture of 6 units of DNA polymerase I (E.coli), 0.6 nanograms of pancreatic DNase I (Worthington), 1 microgram of crystalline BSA in 10 microlitres of 50% v/v

90 glycerol containing 0.05M Tris-HCl, pH 7.5, 0.01M MgCl, and 0.001M dithiothreitol. The mixture was incubated for 2 hours at 15°C, after which high molecular weight DNA was purified by chromatography on G-100 "Sephadex". Figure 13 shows the major bands obtained with DNA

from normal individuals probed with either probe II (Figure 6) or labelled 1.4 kb EcoRl fragment. With each of the 4 enzymes used, EcoRI, HindIII, BallI and Bcll, a single major band of about 4.8, 5.2, 11 100 and 1.7 kb was obtained.

The fact that these restriction fragments had the same length as those observed in the restriction map of clone lambda HIX--1 confirmed that the conditions of Southern blotting were 105 precise enough to detect the factor IX gene in total DNA preparations. This provides the basis for analysis of DNA from the blood of patients with Christmas disease.

(ii) Christmas patients with gene deletions 110 The value of the probes of the invention for the assay of alterations of genes of some patients suffering from Christmas disease has been demostrated as follows. Two patients with severe Christmas disease, who also developed antibodies 115 to factor IX, were selected for study. The DNA from 50 ml of blood was digested separately with EcoRI, HindIII, Bg/II and Bc/I and Southern blots prepared for probing with 32P-nick translated probe II (Figure 6). No specific bands were 120 observed with either patient under conditions where a control digest gave the pattern shown in Figure 13. Similarly no bands were observed in

55

the patients when probe I, III or IV (Figure 6) was substituted for probe II. In order to control for possible mischance of some experimental artefact giving the observed 'negative' signal, a factor IX gene probe (this time pATIXcVII — the cDNA probe) was mixed with an irrelevant autosomal gene probe which was specific for the human Al apolipoprotein (Shoulders and Baralle, Nucl.Acids Res. 10, 4873—4882, 1982). This experiment showed that patient 1 had normal Al

o showed that patient 1 had normal Al apolipoprotein gene, characterised by a 12 kb band in an EcoRI digest, and confirmed that he lacked the 5.5 kb band observed with pATIXcVII and characteristic of the normal factor IX gene. It

15 was concluded that both patients have a sequence of at least 18 kb deleted from their factor IX gene. Two other patients, designated patients 3 and 4, who had also developed antibodies to factor IX gave bands in the normal or abnormal positions on

20 Southern blots with some factor IX gene probes of the invention, but not with others. This suggested that these patients had less extensive deletions of the gene, possibly about 9 kb in length.

These results suggest that diagnosis of

15 haemophiliacs and the heterozygous (carrier)
16 females would be possible in families and this is
17 now under examination. The altered pattern seen
18 in the patient's DNA, whether absence of a band
18 or the presence of a band in an abnormal position,
19 serves as a "disease marker", which can be used
19 to assess for its presence or absence in a

suspected carrier. This same test can be applied to antenatal diagnosis, if DNA from foetal cells are available from an amniocentesis. "Genetic diagnosis" should considerably improve existing

methods of antenatal diagnosis based on the assay of foetal factor IX protein levels, with the added advantage that the test can be carried out earlier in pregnancy. Genetic methods using natural polymorphisms within the factor IX gene as allelic markers should also make 100% carrier

as allelic markers should also make 100% carrier deletion a reality, thereby improving the existing somewhat unsatisfactory methods where probability values are offered to patients.

45 CLAIMS

Recombinant DNA which comprises a cloning vehicle DNA sequence and a DNA sequence foreign to the cloning vehicle, the foreign sequence comprising substantially the following 129-nucleotide sequence (read in rows of 30 across the page):—

ATGTAACATG TAACATTAAG AATGGCAGAT
GCGAGCAGTT TTGTAAAAAT AGTGCTGATA
ACAAGGTGGT TTGCTCCTGT ACTGAGGGAT
55 ATCGACTTGC AGAAAACCAG AAGTCCTGTG

AACCAGCAG

Recombinant DNA which comprises a cloning vehicle DNA sequence and a DNA sequence foreign to the cloning vehicle, the foreign sequence comprising substantially the following 203-nucleotide sequence (read in rows of 30 across the page):—

TGCCATTTCC ATGTGGAAGA GTTTCTGTTT
CACAAACTTC TAAGCTCACC CGTGCTGAGG
65 CTGTTTTTCC TGATGTGGAC TATGTAAATT
CTACTGAAGC TGAAACCATT TTGGATAACA
TCACTCAAAG CACCCAATCA TTTAATGACT
TCACTCGGGT TGTTGGTGGA GAAGATGCCA
AACCAGGTCA ATTCCCTTGG CAG

3. Recombinant DNA which comprises a cloning vehicle DNA sequence and a sequence foreign to the cloning vehicle, the foreign sequence being substantially the same as a sequence occurring in the human factor IX
 genome.

4. Recombinant DNA according to Claim 3 wherein the human factor IX sequence has a length of at least 50 nucleotides.

5. Recombinant DNA according to Claim 3 wherein the length of the human factor IX sequence is from 75 to 27,000 nucleotides.

80

6. Recombinant DNA which comprises a cloning vehicle sequence and a DNA sequence foreign to the cloning vehicle, wherein the foreign sequence includes substantially the whole of an exon sequence of the human factor IX genome.

7. Recombinant DNA which comprises a cloning vehicle sequence and a DNA sequence foreign to the cloning vehicle, wherein the foreign sequence comprises a DNA sequence which is complementary to the human factor IX mRNA.

8. Recombinant DNA according to Claim 3, 4 or 5, wherein the cloning vehicle is a modified pAT153 plasmid prepared by ligating a BamHI
95 and HindIII double digest of pAT153 to a pair of complementary double sticky-ended oligonucleotides having a DNA sequence providing a BamHI restriction residue at one end, a HindIII restriction residue at the other end and a PvulI restriction site in between.

9. Recombinant DNA according to Claim 8 wherein the pair of complementary oligonucleotides are of formula:—

5' GATCCAGCTGA 3'

.

.

3' GTCGACTTCGA 5'

105 10. Recombinant DNA which comprises a cloning vehicle sequence and a DNA sequence

foreign thereto which hybridises to a 247 basepair sequence of bovine factor IX DNA complementary to messenger RNA and indicated in Figure 5 by the arrows at each end thereof.

11. A host transformed with at least one molecule per cell of recombinant DNA claimed in any preceding claim.

12. A host according to Claim 11 in the form of *E.coli*.

10 13. A host according to Claim 11 in the form of mammalian tissue cells.

14. A process of preparing a host transformed with recombinant DNA as claimed in any one of Claims 1 to 7, which process comprises:—

15 (1) synthesising an oligodeoxynucleotide probe having a nucleotide sequence comprising that occurring in bovine factor IX messenger RNA coding for amino acids 70—75 or 348—352 of bovine factor IX and labelling the

20 oligodeoxynucleotide to form a probe;
 (2) preparing complementary DNA to a mixture of bovine RNA.

(3) inserting the complementary DNA in a cloning vehicle to form a mixture of recombinant bovine

25 cDNAs;

(4) transforming a host with said mixture of recombinant bovine cDNAs to form a library of clones and multiplying said clones;

(5) probing the clones with the synthetic
 30 oligodeoxynucleotide probe obtained in step 1 and isolating a resultant recombinant bovine factor IX cDNA-containing clone;

(6) digesting the recombinant bovine factor IX cDNA from said clone with one or more enzymes

to produce a bovine factor IX cDNA molecule containing a shorter sequence of bovine factor IX DNA; and

(7) probing a library of recombinant human genomic DNA in a transformed host with the shorter sequence bovine factor IX cDNA molecule, to hybridise the human genomic DNA to the said recombinant bovine factor IX DNA and isolating the resultant recombinant DNA-transformed host.

15. A process of preparing a host transformed
 45 with recombinant DNA as claimed in Claim 1, 2 or
 7, which process comprises probing a library of clones containing recombinant DNA complementary to human mRNA with a probe comprising a labelled DNA comprising a sequence
 50 complementary to part or all of an exon region of the human factor IX genome.

16. A DNA molecule comprising an at least 15 nucleotide long sequence of part or all of substantially the 129-nucleotide sequence set forth in Claim 1.

17. A DNA molecule comprising an at least 15 nucleotide long sequence of part or all of substantially the 203-nucleotide sequence set forth in Claim 2.

18 A DNA molecule comprising an at least 15 nucleotide long sequence of part only of the DNA sequence of the human factor IX genome.

19. A DNA molecule comprising a sequence of length at least 15 nucleotides substantially the same as a sequence complementary to part or all of that occurring in human factor IX mRNA.

20. A DNA molecule according to any one of Claims 16 to 19 of length at least 50 nucleotides.

21. An artificial DNA molecule comprising a
 sequence substantially the same as a sequence of length at least 15 nucleotides occurring in the human factor IX genome.

22. An artificial DNA molecule according to Claim 21 comprising substantially only exon
75 sequences.

23. A labelled diagnostic probe comprising a DNA molecule having a single-stranded or double-stranded probe sequence of from 15 to 10,000 nucleotides long of DNA sequence defined in Claim 16, 17, 18 or 19 or its complementary sequence.

24. A probe according to Claim 23 having a probe sequence from 20 to 5,000 nucleotides long.

Printed for Her Majesty's Stationery Office by the Courier Press, Learnington Spa, 1984. Published by the Patent Office, 25 Southampton Buildings, London, WC2A 1AY, from which copies may be obtained.

		r.